AMERICAN
JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 17, NUMBER 1 JANUARY, 1949

Jacobian Methods in Thermodynamics

F. H. CRAWFORD
Williams College, Williamstown, Massachusetts

I.

THE usefulness of Jacobians, or functional determinants, in general thermodynamic discussions has not received the emphasis which the power and directness of Jacobian methods warrant. Thus one of the important problems in developing the subject is that of obtaining the so-called "general relations" of thermodynamics. The most important group of these involves the determination of all of those relations involving *first* partial derivatives which must obtain in consequence of the two laws of thermodynamics and certain general and fundamental mathematical theorems. Since it can be shown that for a given physical system any desired derivative can be expressed in terms of an irreducible standard set of first derivatives, the problem becomes that of evolving the most expeditious method for expressing any derivative directly in terms of this standard set. It can be shown that for a general n-variable system (with no special restrictions) this set contains $n(n+1)/2$ derivatives.[1] For a simple homogeneous system under hydrostatic pressure, which we shall have in mind during most of the present discussion, $n=2$, and this standard set contains three *partial derivatives*.

To be more specific we may, after Clausius, express the first and second laws in a combined form as

$$dU=TdS-pdV, \tag{1}$$

where U is the internal energy of the system, S the entropy, and the other symbols have their usual meanings. Our problem, then, is to express any given partial derivative in terms of a standard set by the use of such general mathematical theorems as are available and subject, of course, to Eq. (1). Now these results may be obtained in a great variety of ways, and, in fact, for the system in question are given in a complete form by Bridgman.[2] The usual methods of obtaining these relations are, however, roundabout and tedious, and involve rather cumbersome eliminations. The utility of Jacobians or functional determinants in such cases has received altogether too little attention, despite the fact that Jacobians seem particularly suited for the concise statement of the complex of relationships involved. Several of the older books on thermodynamics[3] express certain derivatives in terms of Jacobians but do not carry the discussion quite far enough for the very real practical advantages to become apparent. More recently, Shaw[4] has given an extremely elegant and elaborate discussion of the hydrostatic case. The fundamental Jacobian theorems used there, however, are of such a form

[1] F. H. Crawford, *Physical Rev.* **72**, 521A (1947). The general results of the n-variable case will be given in more detail elsewhere.

[2] P. W. Bridgman, *A condensed collection of thermodynamic formulas* (Harvard University Press, 1925), pp. 10–15.

[3] See, for example, Bryan, *Thermodynamics* (Leipzig, 1907), p. 22, *et seq.*

[4] A. N. Shaw, *Phil. Trans. Roy. Soc.* **A234**, 299–378 (1935).

as to require considerable dexterity in their manipulation and continual refreshing of one's memory unless they are being used continually.

The method to be discussed here is essentially a special case of Shaw's method but has the practical advantages of being very simple to remember, of needing only *one* intermediate equation obtainable quickly from Eq. (1), and of being capable of ready generalization to systems with $n>2$. There are three steps involved, of which the first two employ Jacobians. They depend for their ease and directness on the use of certain simple and well-known properties of Jacobians. These will be summarized here without proof, though the proofs are simple.[5]

II. Useful Properties of Jacobians

Let us consider, for example, the Jacobians of entropy and volume with respect to the independent variables, x and y. Then we have

$$J(S, V)=\partial(S, V)/\partial(x, y) = \begin{vmatrix} (\partial S/\partial x)_y & (\partial V/\partial x)_y \\ (\partial S/\partial y)_x & (\partial V/\partial y)_x \end{vmatrix}, \quad (2)$$

where the notation $J(S, V)$ will be used wherever the independent variable set is clearly apparent.

(*a*) *Reduction of Jacobians to single derivatives:* Whenever the numerator and denominator of a Jacobian contains a common variable this may formally be "canceled," thus reducing a second-order Jacobian to a single derivative. The canceled variable appears outside the bracket of the partial derivative. Thus,

$$\partial(S, V)/\partial(S, y)=(\partial V/\partial y)_S.$$

Cross cancellation always involves a change in sign. Thus

$$\partial(S, V)/\partial(x, S)=-(\partial V/\partial x)_S.$$

Conversely, of course, we may "expand" any first derivative into a second-order Jacobian.

(*b*) *Change of order of variables* alters the sign of a Jacobian. Thus

$$J(S, V)=-J(V, S)=\partial(V, S)/\partial(y, x), \text{ etc.}$$

(*c*) *A Jacobian does not vanish* unless a functional relation exists between its dependent variables. Thus $J(S, V)\neq 0$ unless $S=S(V)$, which is not true in our homogeneous system of two degrees of freedom. Hence we may divide by and cancel such Jacobians algebraically without ambiguity.

(*d*) *Change of variable theorem:* If, for example, each of the set (S, V) is a function of (x, y) and each of these in turn is a function of a new set (X, Y), then

$$\partial(S, V)/\partial(X, Y)=\frac{(\partial(S, V)/\partial(x, y))}{(\partial(X, Y)/\partial(x, y))}. \quad (3)$$

This is the most general change of variable theorem, and includes as special cases the most useful theorems in partial differentiation. (We shall refer to it in the sequel simply as the Jacobian theorem.)

III. General Method for Obtaining Relations

(*a*) *Step one: Expression of a derivative in terms of no more than four others.* This step consists in expanding any derivative into a Jacobian and then applying the Jacobian theorem, Eq. (3) above. As an example consider $(\partial S/\partial p)_V$. We may write at once

$$(\partial S/\partial p)_V=\partial(S, V)/\partial(p, V) = \frac{(\partial(S, V)/\partial(x, y))}{(\partial(p, V)/\partial(x, y))}=J(S, V)/J(p, V). \quad (4)$$

Now $J(S, V)$ contains four derivatives and $J(p, V)$ two new ones, or six all told. These will in all cases reduce to no more than *four*, however, provided x and y are selected from the set (T, S, p, v) only, since at least one derivative will always vanish and another reduce to unity. Hence, in Eq. (4) we have expressed a partial derivative in terms of the ratio of two independently calculable Jacobians which *at most* can involve four distinct first derivatives. This step, of course, is a purely mathematical transformation and holds whether or not our symbols have any reference to a physical system or not.

(*b*) *Step two: Deduction of Maxwell's equation in Jacobian form.*[6] We must now use some aspect of

[5] See, for example, Margenau and Murphy, *The mathematics of physics and chemistry* (Van Nostrand, 1943), p. 18.

[6] This extremely concise and useful theorem is given by Saha and Srivastava, *A text book of heat* (Allahabad, 1931), p. 408, *et seq.* The above proof is, I believe, new.

the truth contained in the laws of thermodynamics as expressed in Eq. (1),

$$dU = TdS - pdV.$$

The most pertinent way to do this is to utilize the fact the dU is a *perfect* or *total differential* and apply the cross derivative theorem to $TdS - pdV$. This gives us

$$(\partial T/\partial V)_S = -(\partial p/\partial S)_V, \tag{5}$$

which, although a true relation, is in terms of (S, V) as independent variables. To express it in its most general form in terms of, say, (x, y), we have simply tc apply the method of step one to each side of Eq. (5). This gives us for the left side

$$(\partial T/\partial V)_S = \partial(T, S)/\partial(V, S)$$
$$= \frac{(\partial(T, S)/\partial(x, y))}{(\partial(V, S)/\partial(x, y))} = J(T, S)/J(V, S),$$

and for the right side

$$-(dp/dS)_v = -\partial(p, V)/\partial(S, V) = \partial(p, V)/\partial(V, S)$$
$$= \frac{(\partial(p, V)/\partial(x, y))}{(\partial(V, S)/\partial(x, y))} = J(p, V)/J(V, S).$$

Hence Eq. (5) becomes

$$J(T, S)/J(V, S) = J(p, v)/J(V, S)$$

or, since $J(V, s) \neq 0$, we may cancel this Jacobian out of our result and obtain, finally,

$$J(T, S) = J(p, V),$$

or

$$(\partial(T, S)/\partial(x, y)) - (\partial(p, V)/\partial(x, y)) = 0. \tag{6}$$

This, then, is the general Jacobian relation whose special forms are usually known as Maxwell's relations. Since there are six choices of variable sets, if we confine ourselves to (T, S, p, V), we have six forms of Eq. (6). These are not six independent equations but merely six equivalent ways of writing Eq. (5). We may then refer to the appropriate form of Eq. (6) simply as Maxwell's relation.

It is important to note that when (x, y) is a *non-conjugate* set such as (S, V), (T, p), etc., Maxwell's relation always reduces to two derivatives. If we choose a conjugate set, either (T, S) or (p, v), we have an expression involving four derivatives. In all cases, however, no new derivatives appear beyond those arising in step one.

(*c*) *Step three: Combination of results of steps one and two.* We are now in a position to obtain the final result. By step one we have written a given derivative in terms of no more than four other derivatives. From step two we obtain a relation involving two or more of this same group. Hence we may eliminate one of the four derivatives and *express our desired derivative in terms of a standard set of three partial derivatives of the dependent variables with respect to x and y*, provided, of course, that both dependent and independent variables are selected from the set (T, S, p, v).

IV. Specific Examples in Simple Hydrostatic Case

In the above discussion we have left the choice of x and y open. In practice it is convenient to take T and p since they are readily controlled experimentally. To find a given derivative such as $(\partial S/\partial p)_V$ from above, we have, using step one by applying Eq. (4),

$$(\partial S/\partial p)_V = \partial(S, V)/\partial(p, V)$$
$$= (\partial(S, V)/\partial(T, p))/(\partial(p, V)/\partial(T, p))$$
$$= \frac{\begin{vmatrix} (\partial S/\partial T)_p & (\partial V/\partial T)_p \\ (\partial S/\partial p)_T & (\partial V/\partial p)_T \end{vmatrix}}{-(\partial V/\partial T)_p} = \frac{\begin{vmatrix} a & b \\ d & c \end{vmatrix}}{-b}, \tag{7}$$

where the letters a, b, c and d are simply abbreviations for the four derivatives in $J(S, V)$.

Now by step two, Eq. (6), we obtain Maxwell's relation in the form

$$\partial(T, S)/\partial(T, p) - \partial(p, V)/\partial(T, p) = 0,$$

or

$$(\partial S/\partial p)_T = -(\partial V/\partial T)_p, \text{ or simply } d = -b. \tag{8}$$

Since, then, b is more readily measured experimentally than d, we now use Eq. (8) to eliminate d from Eq. (7) and have as our final result:

$$(\partial S/\partial P)_V = (b^2 + ac)/-b.$$

We may now expect to determine all other derivatives which involve only (T, S, p, V) in

terms of the standard set a, b, and c,[7] though in a given case the whole set need not appear. Thus for $(\partial S/\partial V)_p$ we have

$$(\partial S/\partial V)_p = \frac{(\partial(S,p)/\partial(T,p))}{(\partial(V,p)/\partial(T,p))} = \frac{(\partial S/\partial T)_p}{(\partial V/\partial T)_p} = a/b, \text{ etc.}$$

Further, the method is entirely foolproof and, if we apply step one to any of the set a, b, c or d, we obtain simply an identity.

V. Extension to Other Quantities than the Primary Set (T, S, p, V)

We may now remove the earlier restriction to variables of the primary set. There are a number of quantities of special interest which are defined either differentially or integrally in terms of the primary set. These we group in a *secondary* set which will include, among others, the internal energy U, the enthalpy $H(=U+pV)$, the Helmholtz function $A(=U-TS)$, and Gibbs' function $G(=U-TS+pV)$. Consider, then, any derivative such as $(\partial X/\partial Y)_Z$, where X, Y and Z are chosen in any way whatever from these two sets. By step one, $(\partial X/\partial Y)_Z = J(X,Z)/J(Y,Z)$, and will contain only partials with respect to T and p. But these in turn are all expressible in terms of a, b, c and d and hence of a, b and c. The expressions may involve T, S, p and V themselves, but will never involve any derivatives not in the standard set of three. As an example, consider $(\partial U/\partial H)_T$. We have

$$(\partial U/\partial H)_T = J(U,T)/J(H,T) = \frac{(\partial(U,T)/\partial(T,p))}{(\partial(H,T)/\partial(T,p))} = \frac{(\partial U/\partial p)_T}{(\partial H/\partial p)_T}.$$

But $(\partial U/\partial p)_T$ we obtain from Eq. (1) and $(\partial H/\partial p)_T$ from the definition of H (and Eq. (1), again). Thus we have, finally,

$$(\partial U/\partial H)_T = bT + cp/bT - V.$$

[7] Here, of course, $aT = C_p$, $bV = \beta_p$ and $cV = -K_T$, where C_p, β_p and K_T are, respectively, the heat capacity at constant pressure, the coefficient of volume expansion at constant pressure and the isothermal coefficient of compressibility. Their values may be got directly from tables of physical constants.

Now although Q and W, the heat absorbed by and the work done on the system, respectively, are not functions of the independent variables (T,p), we may include them formally in the secondary set on the following understanding. Derivatives such as $(\partial X/\partial Y)_W$ are identical with $(\partial X/\partial Y)_V$, and $(\partial X/\partial Y)_Q$ is the equivalent of $(\partial X/\partial Y)_S$ since no heat is absorbed along an adiabatic. We also replace $(\partial Q/\partial X)_Z$ wherever it occurs formally, by dQ_Z/dX, where Q_Z is the heat function at constant Z. Similarly, $(\partial W/\partial X)_Z$ is to be taken as dW_Z/dX, where W_Z is the work function at constant Z.

If we then add Q and W to our set of primary and secondary quantities, we have ten all told, T, S, p, V, U, H, A, G, Q and W. Any derivative will be expressible as the ratio of two Jacobians $J(X,Z)$ and $J(Y,Z)$. Hence all the conceivable derivatives can be formed if we compute all the Jacobians involving these ten quantities two at a time. There are $10\times 9 = 90$ of these, of which 45 only need be computed since $J(X,Z) = -J(Z,X)$. This establishes the essential mathematical basis behind the tables of Bridgman;[2] he has precisely 90 such entries, and the quantities to be divided to give a desired derivative are simply our Jacobians. In his notation $(\partial X)_Z \equiv J(X,Z)$, $(\partial Y)_Z \equiv J(Y,Z) = \partial(Y,Z)/\partial(T,p)$, etc. Furthermore, we may now compute any entry in the table with Maxwell's relation and the definitions of any secondary quantities needed as our only intermediate equations.

VI. Other Systems

(*a*) *General systems.* The above process need not be confined to the simple hydrostatic system. We may replace Eq. (1) by

$$dU = TdS + Xd\theta, \tag{9}$$

where X is a force variable and θ the conjugate geometric variable. Here, as long as any two of the primary set (T, S, X, θ) may be regarded as the independent variables, we may express any derivative in terms of an appropriate standard set of three.

(*b*) *Restricted systems.* There are many systems to which Eq. (9) applies which are *restricted* in the sense that, for example, X may be a function of, say, T alone. In such cases we may not take

(T, X) as our independent variables but must use another pair such as (T, θ).

Under these conditions the four derivatives which will turn up in step one are always to be found, formally at least, in $J(S, X)$, for S and X are now the dependent variables. But this gives us

$$J(S, X) = \partial(S, X)/\partial(T, \theta)$$
$$= \begin{vmatrix} (\partial S/\partial T)_\theta & (\partial X/\partial T)_\theta \\ (\partial S/\partial \theta)_T & (\partial X/\partial \theta)_T \end{vmatrix}$$
$$= \begin{vmatrix} (\partial S/\partial T)_\theta & dX/dT \\ -dX/dT & 0 \end{vmatrix},$$

since the derivative $(\partial X/\partial \theta)_T = 0$ and Maxwell's relation reduces to $(\partial S/\partial \theta)_T = -dX/dT$. Consequently, we have only *two* independent first derivatives rather than the three of the general case.

The most important example of this restricted case is, of course, that of the two phases of a pure substance under pressure, where p, the equilibrium pressure, is $p(T)$. Other cases are a surface film under tension where the surface tension is a function of T alone, and a simple reversible electric cell (no gas evolution) where the electromotive force depends on T alone. In all such restricted systems we have as many formal Jacobians as in the general case, but they are all simpler.[8]

In practice it is usually quicker and more accurate to determine the desired derivative directly from the appropriate form of Eq. (9) than to calculate the corresponding result for the general hydrostatic case and then to translate the result for the new system.

[8] See reference 2, pp. 18–24.

A Geometrical Introduction to Tensor Analysis for the Physicist

M. J. Walker
The Pennsylvania State College, State College, Pennsylvania

The operations of tensor analysis are not easy to visualize geometrically. This fact is of small concern to the analyst, but is a source of difficulty to most of those who approach tensor analysis for the first time. In particular, the physicist who finds only occasional use for tensor methods is likely to feel the need of a geometrical mnemonic aid for tensor concepts. Furthermore, there are students who feel dissatisfied with their understanding of a concept unless it can be visualized.

This paper shows how many of the concepts of tensor analysis may be visualized geometrically as generalizations of the concepts of vector analysis. The presentation is intended to supplement the analytical development of tensor theory, and therefore makes no pretense of proving anything. The general formulas and concepts of tensor analysis are reduced to the simplest usable special case, and this special case is illustrated geometrically. The general formulas are shown for comparison with the special case formulas. Also, the vocabulary of tensor analysis is retained for the special case of tensors of rank one—that is, vectors—where vector terminology would be more usual.

The subject is developed synthetically, and certain *definitions* are introduced. These definitions are designed for the special case under consideration, and although they will be consistent with the general case, they are not intended to replace the rigorous definitions of the analytical development. Furthermore, it must be emphasized that a geometrical model, like any other analogy, has limitations; the model may have properties that the prototype analytical expression does not have, and conversely.

If one sets out to generalize vector analysis, three possibilities come immediately to the mind:

a. The coordinate axes, along which the components are taken, may intersect at some angle θ, where θ is not 90° as in ordinary vector analysis.
b. The unit of length along the coordinate axes may be different for each axis.
c. The unit of length may vary in size as one proceeds along the axes.

terms of the standard set a, b, and c,[7] though in a given case the whole set need not appear. Thus for $(\partial S/\partial V)_p$ we have

$$(\partial S/\partial V)_p = \frac{(\partial(S,p)/\partial(T,p))}{(\partial(V,p)/\partial(T,p))}$$

$$= \frac{(\partial S/\partial T)_p}{(\partial V/\partial T)_p} = a/b, \text{ etc.}$$

Further, the method is entirely foolproof and, if we apply step one to any of the set a, b, c or d, we obtain simply an identity.

V. Extension to Other Quantities than the Primary Set (T, S, p, V)

We may now remove the earlier restriction to variables of the primary set. There are a number of quantities of special interest which are defined either differentially or integrally in terms of the primary set. These we group in a *secondary* set which will include, among others, the internal energy U, the enthalpy $H(=U+pV)$, the Helmholtz function $A(=U-TS)$, and Gibbs' function $G(=U-TS+pV)$. Consider, then, any derivative such as $(\partial X/\partial Y)_Z$, where X, Y and Z are chosen in any way whatever from these two sets. By step one, $(\partial X/\partial Y)_Z = J(X,Z)/J(Y,Z)$, and will contain only partials with respect to T and p. But these in turn are all expressible in terms of a, b, c and d and hence of a, b and c. The expressions may involve T, S, p and V themselves, but will never involve any derivatives not in the standard set of three. As an example, consider $(\partial U/\partial H)_T$. We have

$$(\partial U/\partial H)_T = J(U,T)/J(H,T)$$

$$= \frac{(\partial(U,T)/\partial(T,p))}{(\partial(H,T)/\partial(T,p))} = \frac{(\partial U/\partial p)_T}{(\partial H/\partial p)_T}.$$

But $(\partial U/\partial p)_T$ we obtain from Eq. (1) and $(\partial H/\partial p)_T$ from the definition of H (and Eq. (1), again). Thus we have, finally,

$$(\partial U/\partial H)_T = bT + cp/bT - V.$$

[7] Here, of course, $aT = C_p$, $bV = \beta_p$ and $cV = -K_T$, where C_p, β_p and K_T are, respectively, the heat capacity at constant pressure, the coefficient of volume expansion at constant pressure and the isothermal coefficient of compressibility. Their values may be got directly from tables of physical constants.

Now although Q and W, the heat absorbed by and the work done on the system, respectively, are not functions of the independent variables (T, p), we may include them formally in the secondary set on the following understanding. Derivatives such as $(\partial X/\partial Y)_W$ are identical with $(\partial X/\partial Y)_V$, and $(\partial X/\partial Y)_Q$ is the equivalent of $(\partial X/\partial Y)_S$ since no heat is absorbed along an adiabatic. We also replace $(\partial Q/\partial X)_Z$ wherever it occurs formally, by dQ_Z/dX, where Q_Z is the heat function at constant Z. Similarly, $(\partial W/\partial X)_Z$ is to be taken as dW_Z/dX, where W_Z is the work function at constant Z.

If we then add Q and W to our set of primary and secondary quantities, we have ten all told, T, S, p, V, U, H, A, G, Q and W. Any derivative will be expressible as the ratio of two Jacobians $J(X,Z)$ and $J(Y,Z)$. Hence all the conceivable derivatives can be formed if we compute all the Jacobians involving these ten quantities two at a time. There are $10\times 9 = 90$ of these, of which 45 only need be computed since $J(X,Z) = -J(Z,X)$. This establishes the essential mathematical basis behind the tables of Bridgman;[2] he has precisely 90 such entries, and the quantities to be divided to give a desired derivative are simply our Jacobians. In his notation $(\partial X)_Z \equiv J(X,Z)$, $(\partial Y)_Z \equiv J(Y,Z) = \partial(Y,Z)/\partial(T,p)$, etc. Furthermore, we may now compute any entry in the table with Maxwell's relation and the definitions of any secondary quantities needed as our only intermediate equations.

VI. Other Systems

(*a*) *General systems.* The above process need not be confined to the simple hydrostatic system. We may replace Eq. (1) by

$$dU = TdS + Xd\theta, \tag{9}$$

where X is a force variable and θ the conjugate geometric variable. Here, as long as any two of the primary set (T, S, X, θ) may be regarded as the independent variables, we may express any derivative in terms of an appropriate standard set of three.

(*b*) *Restricted systems.* There are many systems to which Eq. (9) applies which are *restricted* in the sense that, for example, X may be a function of, say, T alone. In such cases we may not take

(T, X) as our independent variables but must use another pair such as (T, θ).

Under these conditions the four derivatives which will turn up in step one are always to be found, formally at least, in $J(S, X)$, for S and X are now the dependent variables. But this gives us

$$J(S, X) = \partial(S, X)/\partial(T, \theta) = \begin{vmatrix} (\partial S/\partial T)_\theta & (\partial X/\partial T)_\theta \\ (\partial S/\partial \theta)_T & (\partial X/\partial \theta)_T \end{vmatrix} = \begin{vmatrix} (\partial S/\partial T)_\theta & dX/dT \\ -dX/dT & 0 \end{vmatrix},$$

since the derivative $(\partial X/\partial \theta)_T = 0$ and Maxwell's relation reduces to $(\partial S/\partial \theta)_T = -dX/dT$. Consequently, we have only *two* independent first derivatives rather than the three of the general case.

The most important example of this restricted case is, of course, that of the two phases of a pure substance under pressure, where p, the equilibrium pressure, is $p(T)$. Other cases are a surface film under tension where the surface tension is a function of T alone, and a simple reversible electric cell (no gas evolution) where the electromotive force depends on T alone. In all such restricted systems we have as many formal Jacobians as in the general case, but they are all simpler.[8]

In practice it is usually quicker and more accurate to determine the desired derivative directly from the appropriate form of Eq. (9) than to calculate the corresponding result for the general hydrostatic case and then to translate the result for the new system.

[8] See reference 2, pp. 18–24.

A Geometrical Introduction to Tensor Analysis for the Physicist

M. J. Walker
The Pennsylvania State College, State College, Pennsylvania

THE operations of tensor analysis are not easy to visualize geometrically. This fact is of small concern to the analyst, but is a source of difficulty to most of those who approach tensor analysis for the first time. In particular, the physicist who finds only occasional use for tensor methods is likely to feel the need of a geometrical mnemonic aid for tensor concepts. Furthermore, there are students who feel dissatisfied with their understanding of a concept unless it can be visualized.

This paper shows how many of the concepts of tensor analysis may be visualized geometrically as generalizations of the concepts of vector analysis. The presentation is intended to supplement the analytical development of tensor theory, and therefore makes no pretense of proving anything. The general formulas and concepts of tensor analysis are reduced to the simplest usable special case, and this special case is illustrated geometrically. The general formulas are shown for comparison with the special case formulas. Also, the vocabulary of tensor analysis is retained for the special case of tensors of rank one—that is, vectors—where vector terminology would be more usual.

The subject is developed synthetically, and certain *definitions* are introduced. These definitions are designed for the special case under consideration, and although they will be consistent with the general case, they are not intended to replace the rigorous definitions of the analytical development. Furthermore, it must be emphasized that a geometrical model, like any other analogy, has limitations; the model may have properties that the prototype analytical expression does not have, and conversely.

If one sets out to generalize vector analysis, three possibilities come immediately to the mind:

a. The coordinate axes, along which the components are taken, may intersect at some angle θ, where θ is not 90° as in ordinary vector analysis.
b. The unit of length along the coordinate axes may be different for each axis.
c. The unit of length may vary in size as one proceeds along the axes.

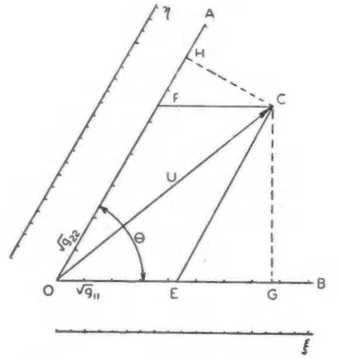

FIG. 1. Contravariant and covariant components of vector $\mathbf{U}$.

It will be found that these three generalizations are sufficient to illustrate many of the concepts of tensor analysis.

I. Covariant and Contravariant Components[1]

For the present, only the first two generalizations will be applied, and the discussion will be restricted to two dimensions. Consider the oblique rectilinear coordinate system AOB (Fig. 1) where the angle $AOB=\theta$. The OB axis is marked off into units of length $(g_{11})^{\frac{1}{2}}$ in terms of the master scale ξ. This is the definition of g_{11}. The OA axis is marked off into units of length $(g_{22})^{\frac{1}{2}}$ in terms of the master scale η. This is the definition of g_{22}. The vector $\mathbf{U}=OC$ is defined equally well by the parallel components OE and OF, or by the perpendicular components OG and OH.

Let

$$\begin{aligned} u^1(g_{11})^{\frac{1}{2}} &\equiv OE, \text{ (the definition of } u^1\text{);}^2 \\ u^2(g_{22})^{\frac{1}{2}} &\equiv OF, \text{ (the definition of } u^2\text{);} \end{aligned} \quad (1)$$

then u^1 and u^2 are the *contravariant components* of $\mathbf{U}$.

Let

$$\begin{aligned} u_1/(g_{11})^{\frac{1}{2}} &\equiv OG, \text{ (the definition of } u_1\text{);} \\ u_2/(g_{22})^{\frac{1}{2}} &\equiv OH, \text{ (the definition of } u_2\text{);} \end{aligned} \quad (2)$$

then u_1 and u_2 are the *covariant components* of $\mathbf{U}$.

[1] The interpretation of covariant and contravariant components given in Sec. I is reported to have been presented by F. Klein in a lecture in 1916, and may be even older.

[2] The superscripts are upper indices, not exponents. Wherever powers are needed they will be written above parentheses. Thus dx^2 should be read "dx superscript two," but $(dx)^2$ should be read "dx squared."

Clearly, the contravariant components u^1 and u^2 are just the number of units of length $(g_{11})^{\frac{1}{2}}$ or $(g_{22})^{\frac{1}{2}}$ required to make up the parallel components OE or OF. The covariant components u_1 and u_2 are just the number of the reciprocal units of length $(g_{11})^{-\frac{1}{2}}$ or $(g_{22})^{-\frac{1}{2}}$ required to make up the perpendicular components OG or OH. Note that u^1 and u^2 and u_1 and u_2 are merely different descriptions of the same vector $\mathbf{U}$. The two descriptions are different when either $\theta\neq 90°$ or $(g_{11})^{\frac{1}{2}}\neq 1$ or $(g_{22})^{\frac{1}{2}}\neq 1$, but become identical in ordinary vector analysis, where $\theta=90°$ and $(g_{11})^{\frac{1}{2}}=(g_{22})^{\frac{1}{2}}=1$. The definition of covariant and contravariant components in this seemingly rather complicated way justifies itself by yielding very symmetrical formulas as the development proceeds. The reason for the terms co- and contra- will be discussed later.

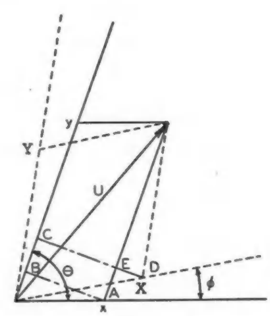

FIG. 2. Rotation of oblique axes: Contravariant components (abscissas) of vector $\mathbf{U}$.

The algebraic manipulation of tensor analysis uses whichever type of component is more convenient at the moment, and changes from one type to the other as conditions require. The process is called *raising the subscript* or *lowering the superscript*, and the necessary formulas are derived below.

From Fig. 1 it is seen that

$$OG=OE+OF\cos\theta,$$
$$u_1/(g_{11})^{\frac{1}{2}}=u^1(g_{11})^{\frac{1}{2}}+u^2(g_{22})^{\frac{1}{2}}\cos\theta,$$
$$u_1=g_{11}u^1+[(g_{11})^{\frac{1}{2}}(g_{22})^{\frac{1}{2}}\cos\theta]u^2,$$

or, in general, $u_i=\sum_j g_{ij}u^j$, where

$$g_{12}=g_{21}=(g_{11})^{\frac{1}{2}}(g_{22})^{\frac{1}{2}}\cos\theta.$$

That is,

$$u_i=g_{ij}u^j, \quad (3)$$

using the sum convention that an index repeated as subscript and superscript indicates summation. Formula (3) permits the change from contravariant components to covariant components. The formula for the inverse change will now be derived.

Solving the following equations for u^1

$$g_{11}u^1+g_{12}u^2=u_1,$$
$$g_{21}u^1+g_{22}u^2=u_2,$$

there is obtained:

$$u^1=(1/g)g_{22}u_1-(1/g)g_{12}u_2, \tag{4}$$

where g is the determinant given by:

$$g\equiv\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix}.$$

If one defines the notation $g^{11}\equiv(1/g)g_{22}$ and $g^{12}\equiv-(1/g)g_{12}$, that is, in general,

$$g^{ij}\equiv(1/g)G^{ij}, \tag{5}$$

where G^{ij} is the cofactor of the element g_{ij} in the determinant g, then Eq. (4) can be written in a form symmetrical with Eq. (3),

$$u^i=g^{ij}u_j. \tag{6}$$

A typical formula of tensor analysis is that for the square of the length of a vector:

$$(\mathbf{U})^2=u_iu^i=u_1u^1+u_2u^2. \tag{7}$$

It is easily shown that Eq. (7) is consistent with the plane geometry of Fig. 1. Similarly, other formulas of tensor analysis may be visualized in this way. In particular, the generalized Pythagorean theorem of tensor analysis,

$$(ds)^2=g_{ij}dx^idx^j, \tag{8}$$

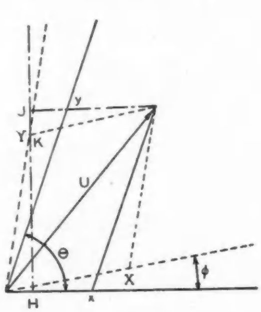

FIG. 3. Rotation of oblique axes: Contravariant components (ordinates) of vector **U**.

may be visualized. In Eq. (8) dx^1 corresponds to u^1, dx^2 to u^2 and ds to $\mathbf{U}$. In expanded form, Eq. (8) becomes:

$$\begin{aligned}(ds)^2&=g_{11}dx^1dx^1+g_{12}dx^1dx^2\\&\qquad+g_{21}dx^2dx^1+g_{22}dx^2dx^2,\\(ds)^2&=((g_{11})^{\frac{1}{2}}dx^1)^2+2[(g_{11})^{\frac{1}{2}}dx^1(g_{22})^{\frac{1}{2}}dx^2]\\&\qquad\times\cos\theta+[(g_{22})^{\frac{1}{2}}dx^2]^2.\end{aligned} \tag{9}$$

II. Fundamental Tensor

The quantities g_{ij} define the metric (i.e., coordinate system) of the space, and are components of the fundamental tensor. The fundamental tensor is of rank two, and thus even in two dimensions has four components. In the simple case under discussion, g_{11} and g_{22} define the length of unit along the axes, and $g_{12}=g_{21}$ define the angle of intersection of the axes. It is in this sense that the fundamental tensor *is* a coordinate system, and thus reference is made to 'differentiating with respect to the fundamental tensor.'

More generally there are the following cases:

$$\begin{array}{ll} g_{ii}=g_{jj}=1,\quad g_{ij}=0: & \text{Euclidean metric;} \\ g_{ij}=\text{constants}: & \text{Cartesian metric;} \\ g_{ij}=\text{functions of } x^i \text{ or } x^j: & \text{Riemann metric.} \end{array} \tag{10}$$

III. Transformation of Coordinates

The physical idea of a vector is that of a directed quantity which retains its direction and magnitude independent of any transformation of coordinate system. The components will, of course, change under transformation. The different behavior of the contravariant and covariant components will now be illustrated under a simple transformation—rotation. First will be needed the partial derivatives, in generalized units, of the transformed coordinates with respect to the old, and inversely. The coordinates themselves, as indicated in Eq. (8), are considered as displacements, and as such are represented by contravariant vectors.

From Figs. 2 and 3, showing rotation of the coordinate axes through a positive counterclockwise angle ϕ to the new dotted position, it is clear that:

$$AB=CD-DE,\quad HJ=HK+KJ. \tag{11}$$

Using capital letters X, Y, Z to denote the new

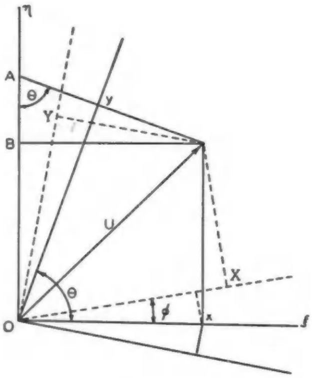

FIG. 4. Rotation of oblique axes: Covariant components of vector **U**.

coordinates and small letters x, y, z to denote the old system, there is obtained from Figs. 2 and 3,

$$x \sin\theta = X \sin(\theta-\phi) - Y \sin\phi,$$
$$y \sin\theta = Y \sin(\theta+\phi) + X \sin\phi.$$

Introducing generalized coordinate units

$$x=(g_{11})^{\frac{1}{2}}x^1, \quad y=(g_{22})^{\frac{1}{2}}x^2$$

as in Eq. (1), there is obtained:

$$\begin{aligned} x^1 \sin\theta &= X^1 \sin(\theta-\phi) - X^2(g_{22})^{\frac{1}{2}}(g_{11})^{-\frac{1}{2}} \sin\phi, \\ x^2 \sin\theta &= X^2 \sin(\theta+\phi) + X^1(g_{11})^{\frac{1}{2}}(g_{22})^{-\frac{1}{2}} \sin\phi. \end{aligned} \tag{12}$$

These yield the partial derivatives:

$$\begin{aligned} \partial x^1/\partial X^1 &= \sin(\theta-\phi)/\sin\theta, \\ \partial x^1/\partial X^2 &= -(g_{22})^{\frac{1}{2}} \sin\phi/(g_{11})^{\frac{1}{2}} \sin\theta, \\ \partial x^2/\partial X^1 &= (g_{11})^{\frac{1}{2}} \sin\phi/(g_{22})^{\frac{1}{2}} \sin\theta, \\ \partial x^2/\partial X^2 &= \sin(\theta+\phi)/\sin\theta. \end{aligned} \tag{13}$$

The inverse derivatives are obtained by setting $\phi=-\phi$ in Eq. (12), and interchanging small and capital letters:

$$\begin{aligned} X^1 \sin\theta &= x^1 \sin(\theta+\phi) + x^2(g_{22})^{\frac{1}{2}}(g_{11})^{-\frac{1}{2}} \sin\phi, \\ X^2 \sin\theta &= x^2 \sin(\theta-\phi) - x^1(g_{11})^{\frac{1}{2}}(g_{22})^{-\frac{1}{2}} \sin\phi. \end{aligned} \tag{14}$$

These yield the partial derivatives:

$$\begin{aligned} \partial X^1/\partial x^1 &= \sin(\theta+\phi)/\sin\theta, \\ \partial X^2/\partial x^1 &= -(g_{11})^{\frac{1}{2}} \sin\phi/(g_{22})^{\frac{1}{2}} \sin\theta, \\ \partial X^1/\partial x^2 &= (g_{22})^{\frac{1}{2}} \sin\phi/(g_{11})^{\frac{1}{2}} \sin\theta, \\ \partial X^2/\partial x^2 &= \sin(\theta-\phi)/\sin\theta. \end{aligned} \tag{15}$$

The transformation formulas for contravariant vectors are the same as Eqs. (12) and (14):

$$\begin{aligned} U^1 &= u^1 \sin(\theta+\phi)/\sin\theta \\ &\qquad + u^2(g_{22})^{\frac{1}{2}} \sin\phi/(g_{11})^{\frac{1}{2}} \sin\theta, \\ U^2 &= -u^1(g_{11})^{\frac{1}{2}} \sin\phi/(g_{22})^{\frac{1}{2}} \sin\theta \\ &\qquad + u^2 \sin(\theta-\phi)/\sin\theta. \end{aligned} \tag{16}$$

By comparing with Eq. (15), these can be written:

$$\begin{aligned} U^1 &= (\partial X^1/\partial x^1)u^1 + (\partial X^1/\partial x^2)u^2, \\ U^2 &= (\partial X^2/\partial x^1)u^1 + (\partial X^2/\partial x^2)u^2. \end{aligned} \tag{17}$$

These are special cases of the general formulas for any transformation of contravariant components of vectors:

$$U^i = (\partial X^i/\partial x^j)u^j. \tag{18}$$

The corresponding formula for covariant components will now be derived. The transformation formulas for perpendicular components on oblique axes are most rapidly obtained by the apparently roundabout procedure of transforming to rectangular axes, rotating through the counterclockwise angle ϕ, and then transforming back to the original oblique axes. From Fig. 4 the transformation from oblique coordinates x, y to rectangular coordinates ξ, η is seen to be $\xi=x$,

$$\eta = OA - AB = y \csc\theta - x \cot\theta. \tag{19}$$

The ξ, η-coordinates are rotated through the angle ϕ:

$$\begin{aligned} \Xi &= \xi \cos\phi + \eta \sin\phi = x \cos\phi \\ &\qquad + y \csc\theta \sin\phi - x \cot\theta \sin\phi, \\ \mathrm{H} &= \eta \cos\phi - \xi \sin\phi = y \csc\theta \cos\phi \\ &\qquad - x \cot\theta \cos\phi - x \sin\phi. \end{aligned} \tag{20}$$

Transforming back to oblique axes by means of the inverse of Eq. (19), there results:

$$\begin{aligned} X \sin\theta &= x \sin(\theta-\phi) + y \sin\phi, \\ Y \sin\theta &= y \sin(\theta+\phi) - x \sin\phi. \end{aligned} \tag{21}$$

The covariant components are introduced using Eq. (2):

$$\begin{aligned} U_1 &= u_1 \sin(\theta-\phi)/\sin\theta \\ &\qquad + u_2(g_{11})^{\frac{1}{2}} \sin\phi/(g_{22})^{\frac{1}{2}} \sin\theta, \\ U_2 &= u_2 \sin(\theta-\phi)/\sin\theta \\ &\qquad - u_1(g_{22})^{\frac{1}{2}} \sin\phi/(g_{11})^{\frac{1}{2}} \sin\theta. \end{aligned} \tag{22}$$

By comparing with Eq. (13) these can be written:

$$\begin{aligned} U_1 &= (\partial x^1/\partial X^1)u_1 + (\partial x^2/\partial X^1)u_2, \\ U_2 &= (\partial x^1/\partial X^2)u_1 + (\partial x^2/\partial X^2)u_2. \end{aligned} \tag{23}$$

These are special cases of the general formulas for any transformation of covariant components of vectors:

$$U_i = (\partial x^j / \partial X^i) u_j. \tag{24}$$

Comparison of Eqs. (18) and (24) shows that the two kinds of components require different coefficients in the transformation formulas. These coefficients provide the analytical criterion of whether a quantity is covariant or contravariant.

The form of the words covariant and contravariant leads one to expect that a covariant vector ought to vary *with* something, while a contravariant vector ought to vary *opposite* to something. Presumably the original inventor of the terms had some such idea in mind; if so, it is no longer obvious what the idea was. If one tries to associate the ideas of *co-* and *contra-* with the new and old coordinates in the transformation formulas (Eqs. (18) and (24)), it would appear that the names ought to be interchanged. If one decides that the prefix *contra-* refers to the fact that the transformation formula (18) is opposite to formula (24), then the prefix *co-* is left without any corresponding meaning. Furthermore, in the sense in which the word covariant is sometimes used in algebra, *both* types of vectors are covariant with respect to transformations of coordinates. It appears necessary to associate each word as a whole with the appropriate transformation formula, and to avoid any connotation suggested by the prefix.

IV. Covariant Derivative

To be considered now is the covariant derivative of a covariant tensor, that is, the derivative expressed in covariant components of a tensor expressed in covariant components. (The contravariant derivative—that is, the derivative expressed in contravariant components—is seldom used.) In the development of this topic the third generalization of vector analysis will be needed, namely, a variation of unit of length along the axis. A familiar example of such variation is semilog graph paper.

In Fig. 5 is shown a portion of semilog paper with the linear scales ξ, x^1, η and the logarithmic scale x^2 where $\eta = \ln x^2$. The element of distance is given by:

$$(ds)^2 = (d\xi)^2 + (d\eta)^2 = (1)^2 dx^1 dx^1 + (1/x^2)^2 dx^2 dx^2.$$

Therefore, $g_{11} = 1$, $g_{22} = (1/x^2)^2$, $g_{12} = g_{21} = 0$. From Eq. (4), $g = (1/x^2)^2$, $g^{11} = 1$, $g^{22} = (x^2)^2$, $g^{12} = g^{21} = 0$. These g_{ij} define the coordinate system. Thus $(g_{11})^{\frac{1}{2}} = 1$ means that the x^1 axis has a constant unit the same length as the master scale unit. The axes are shown to be perpendicular by $g_{12} = g_{21} = 0$, and $(g_{22})^{\frac{1}{2}} = 1/x^2$ shows that the x^2 axis has a variable unit of length $1/x^2$. For example, when $x^2 = 10$, the unit is $\frac{1}{10}$ of the master scale unit. In this case the fundamental tensor may be regarded as a vector of fixed length and direction which moves about the plane. At each point its components define the number of local units Δx^1, Δx^2 which are equal to unit length on the master scale.

Christoffel Symbols: The covariant derivative is expressed in terms of certain combinations of derivatives of the fundamental tensor with respect to the coordinates, called Christoffel symbols of the first and second kind, respectively:

$$[k, ij] = (1/2)(\partial g_{jk}/\partial x^i + \partial g_{ik}/\partial x^j + \partial g_{ij}/\partial x^k), \tag{25}$$

$$\{k, ij\} = g^{kh}[h, ij].$$

It will be shown that these symbols in the covariant derivative formula represent correction terms which are needed to make the covariant derivative independent of the position of the tensor in the coordinate system, that is, to make the covariant derivative acquire one of the most obvious requirements of a tensor. For the two-dimensional Riemann space represented by semi-

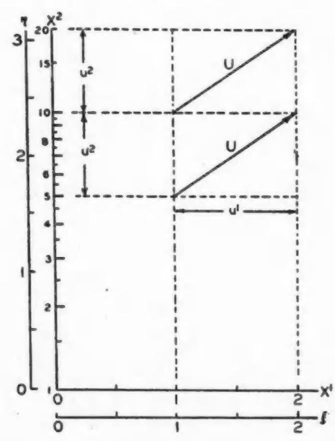

FIG. 5. Movement of constant vector **U** over semilog graph paper.

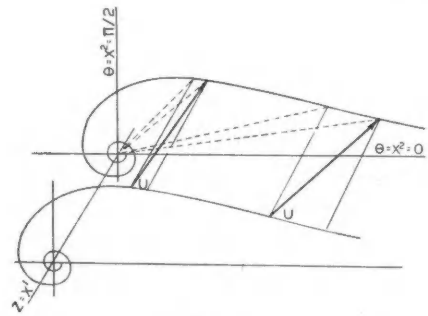

FIG. 6. Two-dimensional Riemann space corresponding to semilog graph paper.

log paper, all Christoffel symbols of the first kind are zero except one:

$$[2,22]=-(1/x^2)^3,$$

and all Christoffel symbols of the second kind are therefore zero except one:

$$\{2,22\}=-(1/x^2). \tag{26}$$

The general formula for covariant differentiation,

$$u_{,k}{}^i=\partial u^i/\partial x^k+u^h\{i,hk\}, \tag{27}$$

leads to the following components of the derivative. (Note that the derivative of a rank one tensor leads to an expression of rank two. This expression of rank two can be proved to be a tensor also.)

$$\begin{aligned} u_{,1}{}^1&=\partial u^1/\partial x^1, & u_{,2}{}^1&=\partial u^1/\partial x^2,\\ u_{,1}{}^2&=\partial u^2/\partial x^1, & u_{,2}{}^2&=\partial u^2/\partial x^2-u^2/x^2. \end{aligned} \tag{28}$$

Referring to Fig. 5, it is seen that a vector of constant direction and magnitude (its position defined, say, by the x^1, x^2 of its midpoint) which moves about the system will suffer no change in its horizontal component u^1 for any change in position x^1 or x^2, no change in its vertical component u^2 for any change in x^1, but will encounter a change in u^2 with a change in x^2. In the latter case $\partial u^2/\partial x^2\neq 0$ although the vector is really constant, and the fictitious derivative $\partial u^2/\partial x^2$ arises because of the changing unit of length. It will be shown that the correction term $-u^2/x^2$ arising from the Christoffel symbol is just enough to subtract off this fictitious change, leaving $u_{,2}{}^2=0$ as one would expect. Thus

$$\begin{aligned}\partial u^2/\partial x^2&\sim\Delta u^2/\Delta x^2\\ &=(x_b{}^2\Delta\eta-x_a{}^2\Delta\eta)/\Delta x^2=\Delta\eta\Delta x^2/\Delta x^2=\Delta\eta,\end{aligned} \tag{29}$$

$$-u^2/x^2=-\Delta\eta/(g_{22})^{\frac{1}{2}}x^2=-\Delta\eta. \tag{30}$$

It was implied above that semilog paper may be regarded as a Riemann space of two dimensions (see Eq. (10)). It can be proved that any Riemann space of n dimensions may be regarded as immersed in a Euclidean space of m dimensions where $m\geqslant n(n+1)/2$. Thus the two-dimensional semilog space may be visualized as a surface in three-dimensional Euclidean space.

Such a surface is shown in cylindrical coordinates in Fig. 6. Its trace on the $z=0$ plane is given by the differential equation:

$$(ds)^2=[r^2+(dr/d\theta)^2](d\theta)^2=(d\theta/\theta)^2. \tag{31}$$

Here z corresponds to x^1, and θ to x^2. On semilog paper the situation is visualized by considering that the size of the unit of x^2 changes with x^2. In Fig. 6 the unit size of θ stays constant, and the distance variation is produced by the curvature of the surface. The case of log-log paper would correspond to a similar surface which also curled into a spiral as one approached the origin along the z axis.

The g_{ij} may thus be viewed from two standpoints: (a) They may be considered as describing a deformation of space where distances are measured in terms of fixed units of length by x^1 and x^2, as in Fig. 6, or (b) they may be considered as describing a variation of unit length of x^1 and x^2 in a flat space with fixed units ξ, η, as in Fig. 5. This duality is the connecting link between tensor analysis considered as a generalization of vector analysis, and tensor analysis considered as a tool for the study of Riemann spaces.

The covariant derivative, Eq. (27), reduces to the ordinary derivative when the Christoffel symbols are zero. This implies that ordinary differentiation may result in the derivative having covariant components even when the two types of component cannot be distinguished. The reason for this behavior may be visualized using Fig. 5.

Consider that a scalar function $\psi(\xi,\eta)$ is plotted along the ζ axis perpendicular to the ξ, η-plane of Fig. 5. The function ψ is then pic-

tured as a surface. One component of gradψ will be $\partial\psi/\partial\eta$, and may be pictured as the slope of the tangent to the surface which is in the plane parallel to the η-axis. Writing this partial derivative in terms of finite increments, and using Eq. (1), there results:

$$\begin{aligned}\partial\psi/\partial\eta &\sim \Delta\psi/\Delta\eta\\ &= \Delta\psi/(g_{22})^{\frac{1}{2}}\Delta x^2 \sim (1/(g_{22})^{\frac{1}{2}})\partial\psi/\partial x^2.\end{aligned}\quad (32)$$

The component of the gradient $\partial\psi/\partial x^2$ is seen to be the coefficient of a reciprocal unit, and in this respect resembles the covariant components of Eq. (2). This picture helps one to visualize why the components of the gradient of a scalar are standard examples of covariant vectors as indicated by the analytical criterion, Eq. (24).

V. Products of Tensors

Visualization of a product of two quantities in general requires more dimensions than the quantities themselves. In the elementary case of $2\times 2=4$, the 2's are visualized as one dimensional, but the 4 is visualized as two dimensional—an area.

The idea of a product of two objects with one component each (say $2\times 2=4$) is established by long familiarity now, but was once a man-made invention requiring as much originality and inventive genius as the invention of the wheel. If one sets out to invent a product of two objects with more than one component each, there are two guiding principles: (a) The product ought to reduce to the ordinary idea of product when the number of components is reduced to one, and (b) The product (or functions of it) ought to be useful for some mathematical or physical purpose.

Consider a pair of objects u^m and v_n with m and n components, respectively. (The use of covariant and contravariant components merely anticipates the fact that greater formal symmetry can be obtained using that combination.) A reasonably general product could be obtained by multiplying each component of u^m by each component of v_n. This procedure yields a product with $m\times n$ components which are conveniently arranged in an array while one considers whether they ought to be added, subtracted, or subjected to some more complicated process. This array itself without any further operations is called the outer or open product.

The outer product of two three-dimensional vectors,

$$\begin{aligned}u^m &= u^1\mathbf{i}+u^2\mathbf{j}+u^3\mathbf{k},\\ v_n &= v_1\mathbf{i}+v_2\mathbf{j}+v_3\mathbf{k},\end{aligned}\quad (33)$$

requires nine dimensions:

$$u^m v_n = \begin{matrix} u^1v_1\mathbf{ii} & u^1v_2\mathbf{ij} & u^1v_3\mathbf{ik}\\ u^2v_1\mathbf{ji} & u^2v_2\mathbf{jj} & u^2v_3\mathbf{jk}\\ u^3v_1\mathbf{ki} & u^3v_2\mathbf{kj} & u^3v_3\mathbf{kk}.\end{matrix}\quad (34)$$

Even for two-dimensional vectors the outer product requires four dimensions:

$$u^m v_n = \begin{matrix} u^1v_1\mathbf{ii} & u^1v_2\mathbf{ij}\\ u^2v_1\mathbf{ji} & u^2v_2\mathbf{jj}\end{matrix}.\quad (35)$$

The physical (spatial) interpretation of these quantities is arbitrary, and the choice has been dictated by the needs of physical theory. In mechanics there arises the equation:

$$\begin{aligned}\text{Energy (a scalar)} &= \text{Force (a vector)}\\ &\quad\times\text{Distance (a vector)}.\end{aligned}\quad (36)$$

Here is a clear need for a type of product of two vectors F and D in which the result is a scalar quantity of magnitude $FD\cos[FD]$ where $[FD]$ is the angle between the vectors F and D. This product—the dot product—is obtained from Eq. (34) by agreeing that $\mathbf{i}\cdot\mathbf{j}=\mathbf{j}\cdot\mathbf{i}=\mathbf{i}\cdot\mathbf{k}=\mathbf{k}\cdot\mathbf{i}=\mathbf{j}\cdot\mathbf{k}=\mathbf{k}\cdot\mathbf{j}=0$, and that $\mathbf{i}\cdot\mathbf{i}=\mathbf{j}\cdot\mathbf{j}=\mathbf{k}\cdot\mathbf{k}=1$. Then,

$$u^m v_n = u^1v_1+u^2v_2+u^3v_3.\quad (37)$$

Also there arise equations of the form:

$$\begin{aligned}\text{Torque (a vector)} &= \text{Force (a vector)}\\ &\quad\times\text{Displacement (a vector)},\end{aligned}\quad (38)$$

where the torque vector is to be perpendicular the plane of F and D, and the magnitude is to be $FD\sin[FD]$. This product—the cross product—is obtained from Eq. (34) by agreeing that $\mathbf{i}\times\mathbf{i}=\mathbf{j}\times\mathbf{j}=\mathbf{k}\times\mathbf{k}=0$, $\mathbf{j}\times\mathbf{k}=-\mathbf{k}\times\mathbf{j}=\mathbf{i}$, $\mathbf{k}\times\mathbf{i}=-\mathbf{i}\times\mathbf{k}=\mathbf{j}$, and $\mathbf{i}\times\mathbf{j}=-\mathbf{j}\times\mathbf{i}=\mathbf{k}$. Then,

$$\begin{aligned}u^m\times v_n &= (u^2v_3-u^3v_2)\mathbf{i}\\ &\quad+(u^3v_1-u^1v_3)\mathbf{j}+(u^1v_2-u^2v_1)\mathbf{k}.\end{aligned}\quad (39)$$

In the Argand diagram for complex numbers, the product is obtained from Eq. (35) by agreeing that $\mathbf{i}=1$ and $\mathbf{j}=\sqrt{(-1)}$. Thus the product of two complex numbers is

$$u^m v_n = (u^1v_1-u^2v_2)+(u^2v_1+u^1v_2)\sqrt{(-1)}.\quad (40)$$

These three examples illustrate the way in which the multidimensional outer product may be reduced to a manageable number of dimensions by conventional agreement in assigning directions to certain components or combinations of components of the outer product.

In tensor analysis, the outer product $u^m v_n$ of two 2-dimensional tensors of rank one has four components and would require four dimensions to visualize it directly:

$$u^m v_n = \begin{matrix} u^1 v_1 & u^1 v_2 \\ u^2 v_1 & u^2 v_2 \end{matrix}. \tag{41}$$

Contraction of Tensors: Somewhat similar to the conventions regarding **i**, **j**, **k** for vectors, there is a process called *contraction* which chooses arbitrarily a certain combination of components of the outer product. Contraction consists of putting one of the covariant indices equal to one of the contravariant indices and performing the sum indicated by the sum convention. In order to contract the outer product shown in Eq. (41) the n is changed to m, thus indicating the sum,

$$u^m v_m = u^1 v_1 + u^2 v_2. \tag{42}$$

The contracted outer product is called the inner product. The process yields a quantity (which can be proved to be a tensor) of rank two less than the original uncontracted tensor. As illustrated in Eqs. (41) and (42), the outer product of two rank one tensors is of rank two, which, when contracted, yields a tensor of rank zero—a scalar. Thus the contraction of the open product of two vectors yields the scalar or dot product of the two vectors. Since any tensor of rank two or greater may be regarded as a product, contraction can be applied to any tensor of rank two or greater, often in several ways. For example, the outer product $u^{ab}v_{cd}$ may be contracted as $u^{ab}v_{ad}$, $u^{ab}v_{cb}$, $u^{ab}v_{ca}$, or $u^{ab}v_{bd}$. The latter contraction $u^{ab}v_{bd}$ corresponds to the multiplication process for determinants and to the product used in ordinary matrix mechanics. It was originally so defined by Cayley for use in the theory of linear transformations. Another example of contraction in a case of particular interest to the physicist occurs in Sec. VI.

Mixed Tensors: The mixed tensor, that is, one with both contravariant and covariant indices, and to which corresponds a mixed transformation formula of the type

$$U_j{}^i = u_l{}^k (\partial X^i/\partial x^k)(\partial x^l/\partial X^j), \tag{43}$$

arises from the product of a covariant and a contravariant tensor. The outer product of two tensors may be shown to be a tensor itself, and thus merits a separate symbol. For example, the outer product illustrated in Eq. (41) may be represented by the single symbol $w_n{}^m = u^m v_n$, where $w_1{}^1 = u^1 v_1$, $w_2{}^1 = u^1 v_2$, etc.

Even in two dimensions this mixed tensor $w_n{}^m$ has four components, and is not easily visualized in three-dimensional space. There is, however, a particular mixed tensor of the second rank which can be visualized. It is the Kronecker delta, $\delta_j{}^i$, where $\delta_j{}^i = 0$ when $i \neq j$ and $\delta_j{}^i = 1$ when $i = j$. Written as an array it has unity on the principal diagonal and zero elsewhere. It is clearly a generalization of the idea of unity, and is related to the unitary matrix. It may be shown that, for any symmetric tensor of rank two, the inner product

$$u_{ij} u^{ik} = \delta_j{}^k. \tag{44}$$

This emphasizes the reciprocal nature of the covariant and contravariant components. If $\delta_j{}^k$ be contracted, $\delta_k{}^k = n$, the number of dimensions of the space. In particular, for the fundamental tensor,

$$g_{ij} g^{ik} = \delta_j{}^k, \tag{45}$$

which is consistent with the "definition" of g_{ij} used in Sec. I.

VI. The Tensor as an Operator

The physicist is frequently introduced to the idea of a tensor as an operator which can produce arbitrary changes of direction in a vector. In an isotropic medium the displacement is in the direction of the applied force, and the equation may be written

$$d^i = K f^i, \tag{46}$$

where d^i and f^i are three-dimensional vectors and K a scalar given bv

$$K = \partial d^i / \partial x^i. \tag{47}$$

In anisotropic mediums, d^i and f^i may not have the same direction, and orthodox vector analysis has no type of product to correspond to such a

physical situation. The equation must now be written

$$d^m = K_i^m f^i, \tag{48}$$

where d^m and f^i are three-dimensional vectors, and K_i^m is a tensor given analagously by the covariant derivative,

$$K_i^m = d_{,i}^m. \tag{49}$$

In expanded form the components of d^m are given by:

$$\begin{aligned} d^1 &= K_1^1 f^1 + K_2^1 f^2 + K_3^1 f^3, \\ d^2 &= K_1^2 f^1 + K_2^2 f^2 + K_3^2 f^3, \\ d^3 &= K_1^3 f^1 + K_2^3 f^2 + K_3^3 f^3. \end{aligned} \tag{50}$$

It is seen that the K's could be determined to correspond to any experimentally occurring situation of d^m and f^i. The tensor K_i^m was introduced to provide an analytical relationship between d^m and f^i for all possible situations of d^m and f^i: As a special case it ought to include the vector cross product where d^m is perpendicular to f^i, that is, $d^m = A^i \times F^i$, or, in terms of components,

$$\begin{aligned} d^1 &= 0 - A^3 f^2 + A^2 f^3, \\ d^2 &= A^3 f^1 + 0 - A^1 f^3, \\ d^3 &= -A^2 f^1 + A^1 f^2 + 0. \end{aligned} \tag{51}$$

By comparison with Eq. (50) it is seen that if K_i^m is a skew-symmetric tensor (that is, if $K_1^1 = K_2^2 = K_3^3 = 0$, and $K_n^m = -K_m^n$) then K_2^3 and A^1, K_3^1 and A^2, and K_1^2 and A^3 may be identified, and Eqs. (50) and (51) become identical. Thus all vector cross products (and hence all axial vectors) are skew-symmetric tensors of the second rank. This occurrence is clearly a special case of contraction of the open product $K_n^m f^i$ of rank three to the inner product $K_i^m f^i$ of rank one—that is, a vector. It can therefore be said that the inner product $K_i^m f^i$, in those cases when it is skew-symmetric, is identical with the cross product of vector analysis; otherwise, it may be regarded as a generalization of the cross product.

Equation (50) also includes the dot product as a special case when all components except those on the principal diagonal are zero.

VII. Dyadics

Dyadics are tensors of the second rank (in three dimensions) written in a special notation originated by Willard Gibbs. In vector analysis the direction of a component is indicated by the unit vectors **i**, **j**, **k** written as a product with the scalar magnitude of the component. There are rules for composition of the **i**, **j**, **k** vectors as in Sec. V. In tensor analysis the function of the **i**, **j**, **k** vectors is performed by the subscripts and superscripts. In dyadic notation the postscripts of second-order tensors are replaced again by *component indices* such as **ii**, **ij**, etc., written as products. There are nine such indices in three-dimensional space, and it is not easy to identify them with unit vectors having fixed directions. The symbols **ii**, **ij**, etc., may be considered as tensor postscripts which have been written in the form of products, and which have their own conventions of composition—an extension of the unit vector method of vector analysis.

VIII. Geodesics

A geodesic may be defined as the path of minimum length between two points, a and b, of a particular space. In two-dimensional Euclidean space $(ds)^2 = g_{ij} dx^i dx^j = (dx^1)^2 + (dx^2)^2$, and

$$s = \int_a^b [(dx^1/ds)^2 + (dx^2/ds)^2]^{\frac{1}{2}} ds. \tag{52}$$

Using a dot to indicate differentiation with respect to s, and introducing the abbreviation symbol ψ, there is obtained:

$$s = \int_a^b [(\dot{x}^1)^2 + (\dot{x}^2)^2]^{\frac{1}{2}} ds \equiv \int_a^b \psi ds.$$

The condition that s shall be an extremum is given by Euler's equations:

$$\partial\psi/\partial x^i - (d/ds)(\partial\psi/\partial\dot{x}^i) = 0. \tag{53}$$

The solution of Eq. (53) is:

$$x^2 = mx^1 + b, \tag{54}$$

the class of all straight lines in the space, as would be expected.

The application of Euler's conditions to lengths in Riemann spaces leads to the general equations for geodesics:

$$d^2x^k/ds^2 + \{k, ij\}(dx^i/ds)(dx^j/ds) = 0. \tag{55}$$

Applied to Euclidean space, the Christoffel sym-

bols are all zero, and Eq. (55) leads to the straight lines given by Eq. (54). For semilog paper, Eq. (55) reduces to two equations:

$$d^2x^1/ds^2=0, \quad d^2x^2/ds^2-(1/x^2)(dx^2/ds)^2=0. \quad (56)$$

The solution is

$$\eta=\ln x^2=mx^1+b. \quad (57)$$

When visualized on Fig. 5 the straight lines $\eta=mx^1+b$ are to be expected; when visualized on Fig. 6 the geodesic curves $\ln x^2=mx^1+b$ are no longer obvious by inspection.

The application of Eq. (55) to the two-dimensional Riemann space represented by the surface of a sphere leads to the well-known result that the geodesics are great circles.

IX. Parallelism

In the vector analysis of the plane the concept of parallel displacement of a vector of constant magnitude leads to the requirement that each component of the moving vector should be constant, that is, the derivatives of each component should be zero. This concept of parallelism must be modified for use on Fig. 5, since the components change as the vector moves about the plane even though its length and direction with respect to the ξ, η-coordinates remains constant.

The generalization required to allow for the changing unit of length leads to the following formula expressing the condition for parallel displacement:

$$du^i/ds+u^i\{i, jk\}(dx^k/ds)=0. \quad (58)$$

The parallel displacement takes place along an arbitrary curve C, and the arc length ds is along that curve.

As an illustration, consider that the curve C is the x^2 axis; Eq. (58) yields two equations:

$$du^1/dx^2=0, \quad (59)$$

$$du^2/dx^2+u^2(-1/x^2)(dx^2/dx^2)=0. \quad (60)$$

Equation (59) expresses the fact that u^1 remains constant as one proceeds up the x^2 axis, which is correct intuitively, and Eq. (60) is the same as Eq. (28) which has already been demonstrated.

The extension of the method to parallelism of vectors of variable magnitude utilizes the idea that the ratio of the components must be held constant.

X. Vector Operators

It remains to complete the generalization of vector analysis by displaying the vector operators written in tensor symbols. The formula for the gradient results from substituting the covariant derivative for the ordinary derivative:

$$\text{grad}\psi=\psi_{,}. \quad (61)$$

The divergence is identified with the contraction of the covariant derivative of a vector. The vector is of rank one, its covariant derivative is of rank two, and the contraction yields the required scalar of rank zero.

$$\text{div}\mathbf{U}=u_{,i}^{\ i}=\partial u^i/\partial x^i+u^h\{i, hi\} =(1/g^{\frac{1}{2}})(\partial/\partial x^i)(u^i g^{\frac{1}{2}}). \quad (62)$$

The Laplacian follows directly by taking the divergence of the gradient:

$$\nabla^2\psi=(1/g^{\frac{1}{2}})(\partial/\partial x^i)g^{\frac{1}{2}}g^{ij}\psi_{,j}. \quad (63)$$

The curl is given by

$$\text{curl}v_i=v_{i,j}-v_{j,i}. \quad (64)$$

If $v_{i,j}=v_{j,i}$ the curl will be zero. This occurs when v_i can be expressed as the gradient of some scalar ψ. Then the order of covariant differentiation is commutable, and the curl reduces to zero.

In conclusion, a warning must be emphasized. As stated in the introduction, this paper is intended only to supplement the conventional treatment of tensor analysis by analytical methods. Any reader whose knowledge of tensors has been gained solely from this paper will have a most inaccurate impression of the subject. It is exactly in those fields which this paper has avoided—the treatment of multidimensional spaces where geometrical visualization is impossible—that the greatest importance and usefulness of tensor analysis lies. It may be said that geometrical concepts have been translated into algebraic manipulations in simple spaces where visualization is possible; these manipulations are then extended to higher spaces where visualization is impossible. The great formal beauty and elegance of the discipline is most apparent in these more sophisticated fields.

Tensor analysis, of course, is independent of any geometrical interpretation whatever, and some mathematicians are justly proud of exhibiting tensor calculus as a pure exercise in algebra. A justification for the sort of discussion attempted in this paper may be found in the words of the mathematician R. Courant:[3] . . . "The attitude of those who consider analysis solely as an abstractly logical, introverted science is not only highly unsuitable for beginners, but endangers the future of the subject; for to pursue mathematical analysis while at the same time turning one's back on its application and on intuition is to condemn it to hopeless atrophy."

[3] R. Courant, *Differential and integral calculus* (Interscience, ed. 2, 1937), Vol. I, p. vi.

XI. Acknowledgment

In preparing this discussion of the fundamentals of tensor analysis the writer has assembled ideas freely from the many currently accessible works on tensors and allied subjects; the point of view governing the choice and arrangement of topics (as described in the introduction) is believed to be novel, and to have important advantages for beginners in the field. The geometrical interpretation of covariant and contravariant components in Sec. I was developed independently by the author, although he recently learned that it had been worked out before. The use of semilog paper as a Riemann space for illustration of some of the ideas of tensor analysis in a simple way is believed to be new.

A New Method in the Teaching of Experimental Physics

F. D. CRUICKSHANK
University of Tasmania, Hobart, Tasmania

THE organization of efficient experimental work in physics in the first and second years of the undergraduate science course presents many well-known problems. The general practice in the universities[1] at present is to follow the orthodox procedure that has come down unchanged in essentials since the institution of teaching laboratories just prior to 1870. With its standardized experiments, laboratory manuals, and general dullness, this method is too familiar to need description. It is remarkable that laboratory work in physics should have remained almost stationary in its outlook and level of organization during a period which has been characterized by such revolutionary advances in so many fields.

In sharp contrast to the early experimental work of the traditional kind, the beginnings of research work come as a great stimulus to the student. There is an absorbing interest in a research problem, and its whole environment is absolutely different from that of the work of the junior laboratories. It may be asked why these stimuli, so powerful and formative in the development of a research student, should remain the exclusive property of the research laboratory and be entirely lacking in the early part of the undergraduate courses. Can these courses be redesigned so that the student is trained from the outset in the systematic scientific method of experimenting? In particular, might it be possible for the instructor in the first- and second-year laboratories to be the director of a research in which the whole of his class is engaged?

In 1938 an experiment was commenced by the writer at the Physics Laboratory of the University of Tasmania to investigate the practicability of a new approach to the problem of teaching experimental physics along lines suggested by these questions. The attempt was made to transfer something of the atmosphere of the research laboratory into the first-year laboratory, the course being planned on the assumption that the function of the first-year laboratory was to lay the foundation of a sound training in the scientific method of investigation. In the discharge of this function those problems were to be investigated which would serve this end best, the coverage obtained over the various sections of the subject

[1] A. Ferguson, *Reports on progress in physics* (Physical Soc., London), **7**, 355 (1940).

being regarded as of lesser importance. To safeguard this matter of coverage, the second point of the plan was the use of demonstration experiments to ensure that the student would become familiar with the general phenomena of the subject over a broad range of topics. The suitable union of these two activities would give the student a wide general acquaintance with physical phenomena and lay the foundation of a training in the scientific method of investigation.

The third point of the plan made the instructor the director of a research group comprising the members of his class. This involved immediately the elimination of the laboratory manual, the instructor being responsible for introducing all problems orally and making the greatest possible use of class discussions in every phase of the work. It was thought that class discussion under his general direction would provide opportunity for putting forward ideas as to the best means of investigating a problem or of surmounting some difficulty which might arise. It would provide ample scope for training in the interpretation of experimental results, in the determination of relations between the physical quantities involved in the measurements made in the experiments, and in the general handling of observational data. The instructor's skill would lie in his ability to draw his group thoroughly into the problem and to inculcate in them the habits of clear thinking and systematic working. To make use of discussions in this way requires, of course, that all members of the class should be engaged on the same problem simultaneously.

The fourth point in the general scheme was the selection of lines of investigation which would provide a program of work continuing over a number of weeks of normal class time, the various phases developing naturally in a connected sequence. This was designed to avoid the scrappy and disjointed character of the orthodox method and to ensure that continuity of thought and inquiry which is characteristic of a normal research problem. The results obtained in the initial stages were so encouraging that a steady transition has been made from the old to the new courses in a series of steps as the evolution of suitable problems and the expansion of apparatus permitted.

It may be of interest to describe the main features of the courses now running at the Laboratory, and to note some of the leading points which have emerged as the result of the experience of the past ten years. In the first year the unit class under the direction of one instructor consists of sixteen students. It is considered that this is the maximum number which can be handled efficiently by one instructor, and although fourteen or even twelve would be more ideal, experience shows that a group of sixteen is workable. A small laboratory sufficient for a group of this size is preferable to a large laboratory in which several such groups might work independently of one another. In the earlier stages of the work here a large laboratory was used, but one of the happier results of the large increase in the number of undergraduates at the close of the war has been the provision of temporary accommodation in which four small laboratories, each about 30 ft by 20 ft in size, have been established in converted army huts. These have proved eminently suitable, and provide accommodation for sixty-four students at any one time, being used in turn by the various sections into which the first-year population is divided. In each group the students work in pairs, the eight pairs being engaged on the same investigation simultaneously. Investigations are selected which lend themselves to extended inquiry, and as the available class time is only four hours per week one line of work may develop from stage to stage over a period of perhaps six or more weeks. This continuity of the investigations promotes very considerable interest and offers scope for vigorous class discussion.

The instructors are drawn in part from the permanent laboratory staff and in part from senior students. The latter are recruited at the end of their second year in many cases, and, in the course of their remaining undergraduate studies and early research training, may act as demonstrators for three or four years in succession. Wherever possible, new recruits work for a period with an experienced instructor before being put in complete charge of a group. Experience has shown that these student instructors are interested in this method of teaching and enter into the work with enthusiasm. Each week all the instructors meet regularly to discuss the lines of investigation in progress, to consider difficulties which may have arisen and methods of meeting

them, and to plan developments and improvements for the course. This ensures a keen critical interest from the teaching side.

The content of the course may vary somewhat from year to year, and even from section to section in the same year. Developmental work in connection with suitable lines of investigation in different fields is constantly in progress, and the content of the courses changes to include the best selection of the investigations which are available. Further, since it is possible to classify the students so that each group of sixteen is fairly homogeneous, those groups which contain the more able students advance at a greater speed and may cover a greater number, and possibly a different selection, of investigations than those groups containing the weaker students. The course thus possesses a flexibility which is able to meet the needs of the various groups so that maximum benefit may be obtained by all.

The major part of one term is devoted to problems connected with dynamics and the properties of matter. In particular, close attention may be given to the study of linear and rotational harmonic vibrations, including the variation of the amplitude of systems vibrating in air, and hence the exponential decay law. Matters such as the moments of inertia of rotating bodies and the elastic properties of springs and torsion wires arise naturally for investigation in the course of these studies. A second term is fully occupied with studies in electricity, covering in a connected investigation such matters as the properties of linear and nonlinear conductors towards steady and alternating applied potential differences; the design of multirange ammeters, voltmeters, and ohmmeters; the measurement of current, potential difference, and resistance using potentiometer and bridge methods; the properties of triode vacuum tubes, the simple triode amplifier, etc. A special aim in this section of the course is to develop an ability in the student to design circuits for various purposes at about the level suggested by the topics just given, for this ability confers a very definite sense of confidence. The third term is mainly taken up with optical studies such as the paraxial properties of spherical lens systems, the elements of interference and diffraction phenomena, and spectra. Within the framework of a course following these general lines it is quite an easy matter to incorporate sufficient problems of the type which would provide practice in measurement or in any other discipline which may be considered necessary, and yet retain the value of the general stimuli associated with it.

The laboratory notebook contains a working diary of all the investigations which are made, together with such notes of the discussions as the student may judge to be valuable. At the end of an investigation it is often useful to hold a group symposium at which various members present reports on the separate sections of the investigation, thus affording a complete review of the whole matter. These reports are discussed and criticized by the students and the instructor, and in some cases the separate reports are put together by two members of the class into the form of a paper. This is finally edited by the instructor, and may be typed and circulated among the students of the year. In addition to this general program, two or three shorter problems may be assigned for unaided investigation after the student has become familiar with the systematic scientific method of working. Such problems are completed by the writing of a paper on the same general lines as the normal report of original research prepared for publication.

In the second year it has been found that twelve students is the maximum number which can be handled efficiently by one instructor on account of the more advanced nature of the work. Present conditions have necessitated groups of sixteen to be under the direction of one demonstrator, but these are definitely too large. Considerably more latitude is permissible in this year as the ability of the students is greater. In addition to more advanced work along the same lines as in the first year, it has been found expedient to undertake problems which, after some general preliminary work, will split into a number of related subsidiary parts for separate investigation by the various pairs of students in the group. Cohesion between the progress of the various branches of the investigation is maintained by frequent discussions.

A number of courses suitable for investigation at the second-year standard have been developed and a selection is made from them to suit the requirements of each group. A study is made of the properties of the ballistic galvanometer leading to

a connected development of its applications to the measurement of capacitance, high resistance, and changes of magnetic flux, and finally to the investigation of the magnetic properties of materials in terms of B–H and I–H curves, etc. The theoretical side of this course provides opportunity for the student to gain a close acquaintance with the differential equation of damped harmonic vibrations in addition to much fundamental electromagnetic theory. A second course is designed to develop the properties of vacuum tubes and their use as amplifiers, pulsing devices, square wave generators, differentiation circuits, etc., providing a sound introduction to electronics. Liberal use is made of a small cathode ray oscillograph unit in this work. In the field of physical optics a study of diffraction is made, developing, for instance, from the single slit through the double and triple slits up to the many-line grating. Wavelength measurements by means of grating and prism spectrometers, and measurements of refractive index and dispersion for a glass, provide scope for experience in precise measurement. On the geometrical optics side a course is being developed for the experimental study of the aberrations of lens systems. A course in elasticity including elastic hysteresis is also available, and this supplies some straightforward work which lends itself to the analysis and handling of the errors of experiment.

The over-all experience of the past ten years is that the introduction of the present method has been very successful. There has been a great increase in the interest shown in the experimental courses, which is evident during the classes and also from the frequent requests to be permitted to extend some investigation outside normal class hours. In the first term of 1947 considerable success was achieved by one group of second-year students who became interested in certain aspects of a fairly stock problem, and following these matters up have produced some work which is apparently original. At the end of the term they presented the results of their work at a symposium at which the whole laboratory staff and the senior students were present. It is hoped to extend this work during a vacation with a view to the possible publication of a paper. This is a very creditable performance for a group of students in the first term of their second year. While this work was in progress a number of these students who came to the laboratory for normal afternoon classes continued their investigations until late in the evening, and in some instances worked right through the night to complete some phase of the work. Though such sessions are not exactly encouraged, for the next day brings its own round of classes, the instances are quoted to show the revolutionary effect which the introduction of the research stimulus has in the junior laboratories.

To illustrate the kind of work which is possible with the methods described a paper is being prepared which gives an account of an experimental investigation of the paraxial properties of a positive thick lens system. This work is undertaken without any background knowledge of lens systems being given, and it is found that at the conclusion of the work the students have gained such an insight into the problem and the methods of working that they can readily undertake the investigation of any other paraxial problem such as a negative lens system, and unravel its properties in a systematic way without aid from the instructor. The work to be described in this paper is presented to the class as a series of problems which develop naturally one from the other, and all the deductions and development of the essential "story" of the lens system come to light during the class discussions. This work is followed by a theoretical course in which the geometrical theory of the lens system is developed, so that the two elements of the scientific method, experimental observations and a tentative hypothesis to explain the experimental behavior, are maintained in their right relation.

This account would not be complete without grateful acknowledgment of the very real help of many persons. First to Professor A. L. McAulay for generously allowing the writer complete freedom of action in the conduct of this work in the first- and second-year classes; then to Mr. B. I. H. Scott and Mr. G. C. Harvey of the staff of this department for enthusiastic support in the more recent phases of the work, and also to many senior students who have given unstinted assistance as demonstrators; and finally to Mr. R. L. Propsting, senior instrument maker at the Laboratory, for very willing and effective cooperation in the production of experimental apparatus, often under very difficult conditions and at short notice.

On Class Room Demonstrations

W. W. Robertson
University of Texas, Austin, Texas

IN these days of many excellent text books, the exact function of an instructor in an elementary physics course is somewhat problematical. But the unique functions of an instructor seem to be threefold: (1) He is the taskmaster for the class, (2) he can give the individual consideration needed by each student (no textbook being sufficiently flexible to meet the needs of all the students), and (3) he can perform demonstrations before the class.

The large classes found in elementary physics courses at the present time reduce the effectiveness of the second function to the point of extinction, through no fault of the instructor. But in the last few years there seems to have been a voluntary tendency on the part of many instructors to give up the third function in favor of a tendency toward fewer and fewer class room demonstrations.

Reasons for the apparent decline in the use of demonstrations are readily found. (1) Apparatus worn out during the war years is still difficult to replace. (2) Apparatus that fell into disuse through lack of repair is probably still in disrepair, since the main effort of a physics department at the present time is quite likely the building up of research apparatus to take care of the large number of graduate students. (3) Some instructors are unwilling to take time from their own research to prepare demonstrations for elementary courses. Less time is required in preparation for a lecture consisting of a formal presentation of familiar material and of the solution of a few problems than for one involving gathering apparatus for workable demonstrations. The effective use of demonstrations also requires strict adherence to a carefully prepared time schedule in order that the material may be covered in the allotted time. (4) The unusually large classes of the last few years present the added problem of obtaining apparatus of sufficient size and simplicity for effective demonstrations.

The answers given to those who question the advisability of spending class room time upon demonstrations are usually that demonstrations (1) relieve the monotony of formal lectures and stimulate interest, (2) impress physical concepts on the minds of the students, (3) illuminate difficult ideas, thus actually saving class room time, and (4) keep the experimental point of view before the students.

In addition to the arguments enumerated above, demonstrations allow an approach to a number of topics in physics which is qualitatively different from that given in textbooks or in a lecture without the use of demonstrations. This latter point will be discussed in considerable detail.

Demonstrations usually serve to illustrate the use of physical principles in the interpretation of natural phenomena or to verify the principles. In these respects no one questions their importance to the students nor grudges them full consideration in lectures and in the laboratory.

An important factor that does not get the attention it deserves in either the lecture or the laboratory is a demonstration of *the manner in which the physical concept is developed in the first place.* To a student in an elementary course, a demonstration of the development of a physical concept is essentially a demonstration of an experimental approach to a scientific problem. The instructor finds the following rather well-defined steps to be taken in a typical demonstration: (1) Illustrating the phenomenon which calls for a new concept, (2) deciding what variables are of importance in controlling the phenomenon, (3) devising an experiment to check the effect of each variable, (4) formulating a generalized statement expressing the relation between the phenomenon and the controlling factors, and (5) verifying the general statement. Since these are the essential steps in the scientific method, lecture demonstrations can be performed in such a manner as to give the student training in the application of the scientific method to a particular scientific problem along the lines outlined above. Such experimentation carries with it a feeling of discovery, of creativeness and of fertility that is often suppressed by the abundance of well-worked-out principles presented to the student.

Although one cannot profitably present all lecture demonstrations from this point of view, there are many which can be so presented. Effort and ingenuity on the part of the instructor are the principal requirements.

The use of simplified demonstration apparatus makes it difficult, in general, to obtain accurate quantitative data, but the effectiveness of the demonstration is not necessarily lessened. Some demonstrations need to give only qualitative answers in order to be useful in illustrating a scientific line of reasoning. The last of the three examples given below is used to illustrate this point. They are given to allow the experienced teacher to judge for himself the possible virtues of this manner of presentation.

Example I

Introduction of the subject of torques to a first year college physics class. The equilibrium of non-concurrent forces is taken up immediately after the completion of the study of equilibrium of concurrent, coplanar forces, and it is, of course, first pointed out that the forces involved may now produce rotation of the object upon which they act. The effectiveness of a force in producing or in tending to produce rotation of an object about some axis is given the name *torque* and the immediate problem is to get a quantitative measure of torque.

The students' own experience can be called upon in determining what factors are involved in expressing the torque due to a force by discussing, for example, the common experience of opening a door. It becomes readily evident that the effectiveness of a force in tending to produce rotation is dependent upon the magnitude of the force, its direction, and its point of application. As many examples as are desired can be added to show that these three factors are the only ones that need to be considered.

The apparatus to be used in the demonstration consists of a meter stick, knife-edge supports, spring balances, and weights. First, the meter stick is suspended at its midpoint upon the knife-edge support so that the weight of the meter stick itself is obviously balanced in so far as its effect in producing rotation is concerned. The forces to be applied to the meter stick are fairly accurate weights suspended by threads looped around the meter stick. A weight is placed at some convenient position near one end of the meter stick and the torque resulting from that weight is balanced by a different weight at an appropriate position near the other end. The clockwise torque is then seen to be balanced by the counterclockwise torque and any quantitative expression for the torques must give the same numerical value for the two above. The values of the factors considered pertinent (force and lever arm) are recorded and the experiment is repeated with different forces and points of application. The "right" combination of force and distance to express torque must be the one that gives the same numerical value for the clockwise and the counterclockwise torques when applied to the observed values. The data can be so chosen that it is obvious that the only simple, yet satisfactory expression is the product of the force and the distance from the point of application of the force to the point of support of the meter stick.

It is easy to show how to handle torques resulting from more than one force in each direction by placing two weights on one side of the support and balancing them by one weight on the other side. It should be emphasized by the demonstrator that the torque due to the single weight can be calculated. The numerical values for the sum of the torques in each direction must again be equal, and it is found that the resultant torque is the sum of the individual torques created by each weight.

If the meter stick is supported at any point other than its midpoint, it is obvious that its own weight is effective in producing rotation. To find out how to account for the stick's weight, it is supported off center and balanced by a single weight on the shorter arm. The next step is to look for the simplest method of expressing the torque resulting from the weight of the meter stick itself, calculating the balancing torque as before. This approach leads naturally to the concept of center of gravity.

The experiment can be repeated with the knife-edge support suspended from a spring balance in order that all forces can be considered. It is then found that regardless of the reference axis chosen for the purpose of calculating the torques, the sum of the torques will always be zero if the meter stick is actually in rotational equilibrium.

Example II

The problem of dry friction. The object here is to discover experimentally the relation between frictional force and normal force. The procedure to be followed is essentially that described above, the essence of the procedure being the reasoned search for pertinent factors which determine frictional forces. A few minutes' experimentation with blocks of different dimensions and of different materials, a spring balance, and a smooth board will give the students sufficient data to show that frictional forces depend upon the nature of the surfaces in contact and upon the weights of the blocks, but that they are essentially independent of the surface areas in contact. Doubling and tripling the weight of a given block by piling weights on it will reveal that the frictional force varies directly with the weight for given surfaces. If the students seem satisfied that all important factors have been considered, a question is then raised about the frictional forces involved in sliding a block up or down an inclined plane. After a brief discussion, the students usually realize that it must be the normal component of the weight that is of importance. This concept is usefully verified by adjusting the slope of the inclined plane until the block slides down at constant velocity and then making the necessary explanations, measurements and calculations.

Example III

Wave motion and the conduction of sound. This example is taken from a portion of an introductory lecture on sound and attempts to show how qualitative experiment may be combined with common experience and common sense to lead to a correct physical concept in a manner familiar to scientific research. The question here is the nature of the disturbance in a medium conducting sound. The idea is first presented that the source of every sound is a vibrating object. A few demonstrations may be given, including that of a card pressed against a rotating toothed wheel and of a jet of air directed against a row of holes in a rotating disk. The problem is then limited to discovering the manner by which energy is transmitted from the vibrating source to the receiver. The first question that presents itself is whether or not a medium is necessary for the transmission of the energy. The experiment of the ringing bell in the bell jar is used to decide between the two alternatives (this demonstration will continue to be used for its pedagogical value even though, as was pointed out recently by Lindsay,[1] the simple explanation that can be given of it in an elementary class is not rigorously correct).

After it is decided that a medium is necessary, the next question presents two additional alternatives: Is the energy transmitted from particle to particle through the medium, or does an individual particle that receives energy from the vibrating source finally impart that energy to the receiver? If the students are reminded that sound is transmitted through liquids and through solids as well as through air, as they probably know on the basis of their own experiences, then they will probably conclude that the energy is transmitted from particle to particle through the medium. Demonstrations such as sound passing through smoke-filled air may then be given to substantiate this conclusion.

The problem is then raised about the nature of the motion of an individual particle in the medium conducting sound energy. A demonstration of the resonant vibration of two tuning forks of the same frequency will show the periodic nature of the disturbance in the space around the source and will indicate that a particle of the medium at that point must be vibrating with the frequency of the source. The finite velocity of sound should indicate that a phase difference exists between adjacent vibrating particles. The facts presented above plus an examination of a model of a simple transverse wave will show that the transmission of sound energy involves wave motion.

Thus, the teaching of the scientific method by lecture demonstrations does not necessarily involve elaborate or novel apparatus. It is rather the manner in which a demonstration is presented and interpreted that determines whether or not it illustrates the scientific approach to a problem. Instruction of this kind can come only from an experienced teacher and researcher, and the instructor who neglects to develop this phase of science teaching is giving up one of the potentially important contributions to his classes.

[1] R. B. Lindsay, *Am. J. Physics* 16, 371 (1948).

A Two-Year Course in Basic Elementary Physics

G. Forman, P. Rudnick, F. G. Slack, and N. Underwood
Vanderbilt University, Nashville, Tennessee

During the past ten years there has evolved at Vanderbilt University a two-year course in physics worthy of description. This course, begun as a compromise with the engineering school, could not have been developed without the support and approval of this group. Prior to the experiment with the two-year sequence the regular sophomore four-hour course was given. The electrical and mechanical engineering departments desired to start courses in the sophomore year which they felt should be preceded by physics. The physics department desired freshman mathematics as a prerequisite and calculus as a corequisite and thus did not wish to move its course into the freshman year. The department also felt the need of some modern physics in the course and desired more than four year-hours to cover elementary physics including this. Thus, a compromise arrangement, wherein a three-hour per week course extended over two years beginning in the freshman year, was begun on an experimental basis. This increased the credit content by 50 percent to the satisfaction of the physics department, introduced mechanics and electricity in the freshman year to the satisfaction of the engineering departments and, requiring only two lectures and one laboratory period per week, fitted into the freshman-sophomore programs satisfactorily. Actually the course gives approximately 120 lecture periods as compared with 60 on the older system and it retains 30 three-hour experimental laboratory periods spread over the two years. In place of the one-hour weekly quiz section it substitutes a biweekly three-hour practice problem period which alternates with the laboratory, and running for two years, provides thirty of these sessions. This arrangement has proved very satisfactory to all concerned.

Spreading the course over the two-year period appears to have advantages which more than compensate for such obvious difficulties as too little concentration and waning interest. Some of the advantages are the correlation of mathematics and its application in physics, and of physics and its application in engineering subjects. The instructor, forced to begin the course in a very elementary fashion, has the feeling of not being rushed. Subsequently he finds the pace can be increased so that in the sophomore year material can be covered at a considerably accelerated pace and on a higher level. The two-year period gives opportunity for "soaking in" which appears far superior to that afforded by short concentrated courses.

The text used is *Analytical experimental physics* by Lemon and Ference. Six quarters of the course are divided approximately as follows:

> The first two quarters are devoted to mechanics and are used to give as full an understanding of basic principles and terminology as possible. Vibration and periodic motion are included in such manner as to reduce the time that would otherwise have to be spent on sound. The third quarter covers d.c. electricity and the fourth a.c. electricity and introductory electronics. The latter subject, taken in the first quarter of the sophomore year, coincides with and is strengthened by the engineering courses in mechanics and electricity. The fifth quarter covers sound and heat, including radiation and Planck's quantum equation. The final quarter is divided between optics and atomic and nuclear structure approached from the viewpoint of spectra. The course includes modern physics as a sixth topic added to the conventional five fields of mechanics, electricity, and magnetism, sound, light, and heat. Approximately one sixth of the course is devoted to modern material discussed in several different quarters under the headings of electronics and radiation. About one half of the last quarter is actually labeled modern physics.

The use of the three-hour laboratory period every other week for working problems is an unusual feature. With properly selected instructors and not more than 20 students per section both learning and teaching proceed effectively. No new subject matter is introduced in these periods. Occasionally, demonstration experiments are performed, data taken, individual computations made and the results discussed. Atwood's machine, e/m, cathode-ray tube, Schilling sound interference, viscosity measurement, ratio of specific heats, diffraction grating and Geiger counter absorption measurements are typical of

such experiments in various fields. Usually, illustrative problems are assigned for work by the class. The subject matter of these problems is correlated with the lectures. Projectiles, electrical circuits, thermometry, and image formation by lenses are examples of fields covered. The students are encouraged to proceed as far as possible with solutions to problems and derivations in the general field. They are told that they should acquire a feeling of satisfaction and a mastery of the principles under discussion by freely asking questions and assistance of the instructor. This aim sounds idealistic but to most instructors it is a challenge; and the students' response is generally to take advantage of the opportunity for supervised study. Frequent quizzes are given during the regular lecture periods. Two-hour examinations are given quarterly in accord with general practice in the college.

In summary it is found that this two-year (three credit hours per week) course in elementary physics gives adequate training to satisfy the engineering school and can be fitted into schedules without difficulty. It is a great improvement over the standard sophomore physics course. It does not pretend to achieve the depth and coverage of the four- to six-hour courses running for two years at some of our larger engineering schools, but it is a very reasonable compromise. The course has proved so satisfactory that all science majors including N.R.O.T.C. and premedical students are advised to take this sequence. The equivalent of this course is made a prerequisite to senior college courses in physics.

The Delusion of the Scientific Method

HAYM KRUGLAK
University of Minnesota, Minneapolis, Minnesota

I. The Scientific Method as a Blind Spot

EVERYBODY talks about *the* scientific method! It may replace the weather as a common topic of conversation among educators. In the periodical literature of the past two decades there has been an ever-increasing stream of articles on and references to the scientific method. Frequently, these essays and studies are speculative, authoritarian, metaphysical, or generally inconclusive—the very antithesis of the method under discussion. The writers and speakers refer to the *steps* of the scientific method in the naïve, if sincere, faith that these constitute the invariable sequence of a scientific inquiry.

The history of science shows that the acceptance of a belief common to a particular culture has often acted as a blind spot in the quest for truth. Thus, the Greek belief that the circle was the most perfect curve was accepted even by Copernicus.[1] It was the persistent genius of Kepler that finally broke the two thousand year old stranglehold on astronomy. Has the rigid formulation of the scientific method become one of the blind spots of western culture?

II. What the Scientific Method Is Not

The so-called scientific method is not a philosophy of life, nor a procedure by which every problem of living can be solved. The scientific method is not a formula which the scientist applies in the laboratory. The scientific method is no guarantee that the user, however expert, will not make mistakes—will not draw the wrong conclusions. "The very method of science, the way in which it defines a fact and its essential presuppositions, is not subject to scientific proof."[2]

If the scientific method were a definite, well known series of steps, then it would be only necessary for anybody to memorize and drill those steps in order to solve any problem. Also, with a given set of data, everyone familiar with the recipe should arrive at the same conclusion. The history of science shows, however, that several great scientists can work with a given set

[1] Nicholas Copernicus, *De revolutionibus celestium*, quoted in Shapley and Howarth, *A source book in astronomy* (McGraw-Hill, 1929), pp. 1–3.

[2] The Harvard Report, *General education in a free society* (Harvard Univ. Press, 1945), p. 65.

of data and fail to see the underlying principle. Then, some other investigator examines the very same evidence and finds the basic relationship. The law of radiation is a good illustration of the above point. Both Wien and Rayleigh derived formulas for the distribution of radiant energy in the spectrum of a blackbody but neither was successful in fitting all the experimental evidence into a single relationship. Planck attacked the same problem and derived the law of radiation associated with his name, which was in good agreement with the observed facts. Are we to conclude that Planck was a better scientist than either Wien or Rayleigh? Does it follow that Planck knew the scientific method, whereas Wien and Rayleigh were ignorant of it?

III. What Is the Scientific Method?

There are many scientific methods. The nature of the problem determines the particular method to be used.

> It should be an important aim of general instruction in science . . . to give a clear appreciation of the hierarchy of nature and its reflection in the hierarchy of science . . . to convey a most important generalization: that all modes of inquiry must be adapted to the material under consideration and to available methods of approach . . .[3]

Scientific methods depend also on political and economic pressures, religious beliefs, available tools and instruments. "Neither inquiry nor the most abstract set of symbols can escape from the cultural matrix in which they live, move and have their meaning."[4] Since the scientific methods are a function of a given culture, then it is important to recognize their *dynamic* aspect.[5]

It is impossible to predict what will be the exact procedure for solving a given problem until that particular problem has been solved. It is only then that the scientist, if questioned, retraces his steps and describes the procedure, omitting perhaps a good many important details which, have not reached the conscious level. Thus, describing the particular scientific method, as a method, is always an *a posteriori* process.

[3] The Harvard Report, *op. cit.*, pp. 154–5.
[4] Dewey, *Logic; The theory of inquiry* (Holt, 1938), p. 20.
[5] For a penetrating analysis of the changing and evolving science, see: Cameron, "The current transition in the conception of science," *Science*, **107**, 553–558 (1948).

It is generally conceded that there is no satisfactory definition of the scientific method. It is true that habit patterns become established in scientific research as elsewhere, but some of the most significant contributions were made by those who were courageous enough to leave the static framework of conventional methodology. This point of view was clearly recognized by Felix Kaufmann when he stated that "the most momentous scientific achievements were not realized by a procedure in conformity with established rules, but rather in violation of the rules, and these achievements have led to the establishment of new rules superseding the old. . . . If Planck, Einstein, and Bohr had observed the rules of physical inquiry generally adopted at the beginning of the twentieth century, they would not have become the founders of quantum mechanics."[6]

The majority of scientists would probably agree with P. W. Bridgman's colloquial description: "The scientific method, as far as it is a method, is nothing more than doing one's damnedest with one's mind, no holds barred."[7] Translating the above statement into the language of operationism: the scientific method is a set of operations used in finding a solution to a problematic situation which arises out of a given set of empirical facts.

IV. The Teaching of the Scientific Method in Colleges

The lockstep determinism of the 19th century still dominates the teaching of science in the elementary college courses of the United States. There appear to be at least three presuppositions: First, the scientific method is a closed set of rules.

> "Pour un observateur superficiel, la verité scientifique est hors des atteintes du doute; la logique de la science est infallible et, si les savants se trompent quelquefois c'est pour avoir méconnu les règles."[8]

Second, exposure to and experience with scientific material results in a transfer of training to successful solution of problems in other fields.

[6] Kaufmann, "The nature of scientific method," *Social Research*, **12**, 469 (1945).
[7] Bridgman, "The prospect for intelligence," *Yale Review* **34**, 450 (1945).
[8] Poincaré, *La science et l'hypothèse* (Flammarion, Paris, 1902), p. 1.

"Modern science, as training the mind to an exact and impartial analysis of facts, is an education specially fitted to promote sound citizenship."[9]

Third, a knowledge of facts is good in and of itself.

"In high schools and college, laboratory and microscopic observations are carried on as if the accumulation of observed facts and the acquisition of skill in manipulation were educational ends in themselves."[10]

The dynamic nature of science has discredited the first assumption. Educational psychology is vehemently denying the validity of the other two. And yet the majority of the college physical science courses are taught in much the same fashion as they were fifty years ago. The only complaint that one hears from instructors is that there is too much material to cover in the alloted time.

Paradoxically, the symptoms of the *scientific-method blind spot* are very prominent where one expects them least—among teachers of physical sciences in colleges. For instance, the author of an excellent and justifiably popular college textbook of physical science transmits the shibboleth of our times in the following terms: "The scientific method contains nothing mysterious, nothing complex. Reduced to simple terms, it may be divided into three steps: (1) observation, (2) generalization from the observed facts, and (3) checking the generalization by further observations."[11]

In a laboratory manual for a general education course in the physical sciences, it is admitted that the first forty experiments "are designed primarily to demonstrate some of the basic principles of the physical sciences" and, that in such experiments, "the conclusion drawn will generally coincide with the object of the experiment."[12] Here again the word experiment is poorly chosen to describe the type of activity that the student carries out. But at least the author does not claim that the student will acquire the habit of inquiry and scientific training.

However, in the section entitled *scientific method experiments*, the author observes quite correctly that "In original scientific research the answers are not known and the conclusions to be obtained cannot be stated in advance. One makes hypotheses and sets up theories about probable answers and conclusions, but only correct interpretation of the experimental evidence enables one to arrive at facts."[13]

But from there on he gets caught in the pseudo-Dewey trap of *the* scientific method and proceeds to outline its steps in accordance with the interpretation of the self-appointed apostles of Mr. Dewey, and applies these steps to the problem of purchasing a fountain pen.

The scientific or logical methods of solving a problem may be outlined as involving the following steps:

1. Make a statement of the problem.
2. Formulate a working hypothesis.
3. Collect data.
4. Test the data against the hypothesis.
5. If sufficient data agree with the hypothesis, then the hypothesis becomes a theory.
6. More data are accumulated and if, within the limits of experimental error, all the data agree and none disagree, the fact is established.
7. If at any time the data disagree with the hypothesis, theory, or fact, a new hypothesis must be set up and the process repeated.[13]

It is true that one can recognize the similarity between the above steps and Dewey's five steps of reflective thinking.[14] However, Dewey surely did not expect those steps to become the pedagogical catechism of science teaching, as it has unfortunately happened. He foresaw the dangers of an inflexible method. "It is, however, a common assumption that unless the pupil from the outset *consciously recognizes and explicity states* the method logically implied in the result he is to reach, he will have *no* method, and his mind will work confusedly or anarchically . . . But because teachers find that the things which they themselves best understand are marked off and defined in clear-cut ways, our schoolrooms are pervaded with the superstition that children are to begin with already crystallized formulae of method."[15]

Consider a typical experiment in a laboratory manual of elementary college physics: Boyle's

[9] Pearson, *The grammar of science; Part I—Physical* (Adam and Charles Black, Ed. 3, 1911) p. 9.

[10] Dewey, *How we think* (Heath, 1910), p. 192.

[11] Krauskopf, *Fundamentals of physical science* (McGraw-Hill, Ed. 2, 1948), p. 86.

[12] Haun, *Laboratory manual to accompany foundations of the physical science courses* (University of Kansas Book Store, 1947), pp. 7–8.

[13] See reference 12, p. 90.

[14] See reference 10, p. 72.

[15] See reference 10, pp. 113–114.

Law. The law is stated, the apparatus is described, the necessary formulas are derived, a table for entering readings is provided, and the necessary precaution for uniform temperature is given. Shades of Robert Boyle! The chances are good that the lecturer has already talked about the law in lecture and perhaps demonstrated it; that the student has had physics in high school and performed the experiment there; he is likely to have had chemistry in college and high school and to have heard about it there. When the student is now asked to do an *experiment* on Boyle's Law, the spirit of scientific inquiry degenerates into a habit of collecting the same data that has been collected by millions of other bored students. This is "inventory taking of Nature" for the sake of taking inventory.

It is perhaps not difficult to understand how the teaching techniques in a given period are a reflection of the practices of previous generations: imitation is often an escape from ignorance. But what cannot be condoned is the failure of a science instructor to understand the dynamic character of the field. The educational result is not only that the student does not see the forest on account of the trees, but that the trees themselves appear to be petrified.

V. Is It Possible to Teach the Scientific Method?

In the above discussion it was brought out that (1) the current views of the scientific method are probably crude over-simplifications of the actual complexities of a scientific inquiry, (2) the presentation of the scientific method as a series of steps in a never changing sequence, is most unrealistic and leads to pedagogical sterility.

What about induction, deduction, controlled experiment, mathematical formulation? These are the tricks of the trade that increase the probability of solving the problem. They are important elements in all investigations of physical reality, but their relationship to scientific inquiry is quite complex, and often defies analysis. There is some evidence that these special techniques can be taught successfully in public schools.[16,17] An understanding of these special tricks is a necessary but not sufficient condition for the understanding of the nature of scientific inquiry. Consider the case of Galen who lived in the second century A.D. and was one of the first experimental physiologists. He was the first to demonstrate the function of ureters by performing a *controlled* experiment on a living animal.[18] Yet, when he tried to give an explanation of magnetism, he used *logic* instead of empirical facts, and became involved in a net of fantastic statements.

It is generally agreed that an understanding of the development of scientific inquiry and some skills in the use of its methods are important to every individual, if he is to make a satisfactory adjustment to our present culture. However, it is nothing short of a delusion to claim that the present teaching of science is anything but a distorted image of the scientific method. Fortunately, there are several straws in the wind which indicate that college science teachers are becoming aware of the discrepancy between what science is and how it is taught.

VI. Recent Trends in the Teaching of College Science

Some of the factors which might contribute to the improvement of college science teaching are: (1) a growing volume of critical appraisal of the philosophy, methods and techniques of teaching, (2) designs and tryouts of new courses for general education curricula, (3) a recognition that college teachers need special pedagogical preparation, (4) the extension of the scientific approach to the teaching of science.

One of the most blunt and outspoken critics of the present state of science education is a science teacher himself. "It has surprised and startled me that physicists should have failed so badly in making physics contribute to the general education of nonscientists. . . . What we need at this point is more action and less talk."[19] Philipp Frank maintains that science instruction is far from satisfactory even in the training of scientists. "The result of conventional science teaching has not been a critically minded type of scientist, but

[16] Curtis, "Teaching scientific methods," *School Sci. and Math.* **34**, 816–819 (1934).

[17] Sharpe, "Why Not Use Control Experiments?," *Science Educ.* **12**, 19–22 (1938).

[18] Galen, *On the natural faculties*, A. J. Brock (Harvard Univ. Press, 1947), pp. 55–61.

[19] Dodge, "New frontiers," *Am. J. Physics*, **16**, 209–216 (1948).

just the opposite . . . the present type of science instruction does not enable the student to form even the faintest judgment about the interpretation of recent physics as a part of the new world picture. This failure prevents the science graduate from playing in our cultural and public life the great part that is assigned to him by the ever mounting technical importance of science to human society."[20]

Many science teachers are aware of the external and internal criticisms leveled at the traditional methods and courses. The College Physics Teachers Colloquium, at the University of Iowa, and the Conant-Taylor-French Conference on the Teaching of General Science, at Princeton University, were doubtless helpful in focussing attention on current issues, in suggesting possible solutions, and in discussing new approaches to science teaching. However, the ultimate answer lies with the rank-and-file science teacher. He must be willing to explore continuously for new and better methods; he must be prepared to change quickly his practices in conformity with the dynamics of the times. That many science teachers throughout the country are trying out new schemes is a hopeful sign. One of the promising and much talked about designs for a general education science course has been outlined by President Conant of Harvard University. The proposed course on the tactics and strategy of science would use the historical approach: "The stumbling way in which even the ablest of the early scientists had to fight through thickets of erroneous observations, misleading generalizations, inadequate formulation, and unconscious prejudice is the story which it seems to me needs telling."[21] Another approach is exemplified by the course offered at Colgate University, *Problems in Natural Science.* Several topics are selected out of various areas in the physical sciences. The student makes a case study of "how the problem got there, how it was attacked, and what major elements contributed to its solution, and what it means to society."[22]

[20] Frank, "The place of the philosophy of science in the curriculum of the physics teacher," *Am. J. Physics*, **15**, 202 (1947).

[21] Conant, *On Understanding Science* (Yale Univ. Press, 1947), p. 15.

[22] French, "The need for a new approach to science teaching," *The Wiley Bulletin*, December, 1947.

But the inevitable question which presents itself is: *How many teachers are prepared to present science from an historical point of view?* How many teachers are aware of the fact that the scientific method was used before Galileo? How many scientists know that Galileo did *not* drop stones from the leaning tower of Pisa? Until more qualified teachers become available, the majority of colleges will not be ready to offer a course on the tactics and strategy of science. However, there is a crying need for such a course in every graduate school. The history of scientific thought, the development of investigations, and the evolution of ideas and methods should prove invaluable to every budding scientist. A seminar, course, or colloquium on the history and philosophy of science would make a genuine contribution to the background of the present and future science teacher. Such a program could be handled on a departmental or interdepartmental basis.

Unfortunately, the resistance to any change in pedagogical practice increases rapidly as one ascends the academic ladder. The attention which some graduate schools are giving to the problem of college teacher training is in itself a very healthy educational symptom. But it is difficult to predict how many years will pass before teachers with the contemplated special training will go out into the field. In the meantime there is no reason why a more limited training program should not be instituted on a divisional or departmental basis.

One of the weak links in the teaching of college science is the untrained, unprepared, and often unsupervised teaching assistant. The least experienced assistants are usually assigned to the teaching of laboratory classes. Yet, the laboratory is supposed to be the workshop of the scientist—the very place where the student is to learn most about the scientific method. In-service training courses can and should provide the much needed opportunity for demonstrations and discussion related to problems that arise in the laboratory and the quiz section.

A few years ago the writer had the opportunity of organizing and conducting a training course for about forty inexperienced Signal Corps instructors. During the weekly meetings the aims of the program were clarified, effective teaching techniques demonstrated, evaluation instruments dis-

cussed and planned, and practical classroom situations analyzed. The morale and performance of the staff and the students improved markedly in a few months.

The general education science program at the University of Minnesota involves relatively few students and a correspondingly small teaching staff. The supervision and training program is very informal, but intensive. Frequent conferences among instructors, careful coordination of lecture and laboratory work, discussions with the assistants of each laboratory exercise, frequent visits to the laboratory, interviews with students, and the continuous revision of the material, all contribute to make the courses fluid and functional.

Every attempt is made to avoid the stereotyped "experiments" in the laboratory for the general education science courses. Conventional laboratory work does not begin to tap a small fraction of the student's resourcefulness. What better way is there to teach that the scientific method is not a superhighway than to place the student in a situation where he will experience the same failures, make the same mistakes, suffer the same accidents, and explore the same blind alleys as the research scientist in the course of his daily work?

For example, the first laboratory exercise is to construct a simple instrument for locating a star. The student is given a drawing of the Horizon system in astronomy. There are available mailing tubes, knives, cardboard, protractors, rulers, etc. The student has to devise a method of constructing a simple instrument. He submits his plan to the instructor for approval. If the latter approves it, the student may proceed with the construction of the instrument. When the instrument is completed, the student locates the position of the two "stars" pasted on the laboratory walls. The instructor checks the results to be sure that the student knows how to use his instrument.

Not all of the exercises are of this nature. Some, of course, are designed to make textbook material more meaningful; others to develop certain skills. But in all exercises, the number of directions given is kept at a minimum. At no time is the student forced to channel his efforts within the narrow limits set by a number of prescribed steps. When the instructor sees the student pursuing a more complicated procedure than necessary, he may let him go ahead with it; afterwards, he may suggest a shorter route.

Most of the students are so bewildered by this unusual laboratory procedure that it takes them a considerable time to recover from the initial shock. They soon realize that it is either sink-or-swim, and the majority of them start 'swimming.' However, the lifeguard—the instructor—is on hand to rescue the few who start going down. Eventually everybody learns to 'swim' more or less well.

VII. The Teacher's Responsibility

It is generally recognized that the teacher is the key element in the dynamic approach to science instruction. Recognizing this, many universities are beginning to experiment with programs for the preparation of college teachers. As time goes on, it becomes increasingly clear that under present trends in American education, a man distinguished for his scholarship is not necessarily the best teacher. As a recent article put it, "time, numbers of students, and the pressure of unmet or ill-met needs now confront graduate education with problems that cannot and should not be shunted along that familiar buck-passing line that leads ultimately to the kindergarten, preschool, and home. . . . The problem of graduate training of college teachers is immediately pertinent because more than half of America's Ph.D.'s go into teaching."[23]

Perhaps the greatest handicap to the improvement of science teaching is the failure of science teachers to agree on an objective, systematic procedure for attacking the problems of science education. Here, the blind spot covers the entire educational retina. "Where curriculum committees have struggled to bring greater unity into the curricula of liberal colleges, the scientists have too often been the hold-outs. Or they have come in with mental reservations, unwilling, it almost seems, to carry their vaunted experimental methods of science over the experimental problems of *teaching* science."[23]

President Dodds of Princeton University speaks in the same vein: ". . . the neglect to experiment

[23] Blegen, "The graduate schools and the education of college teachers," *Educ. Record*, **29**, 12–13 (1948).

more in search of factual evidence is surprising in a profession staffed by men whose lives are largely dedicated to the collection and transmission of systematic knowledge."[24] It is amazing and disturbing to find in eminent scientists and science educators a naïve, almost pre-Galilean approach to the pedagogical aspects of modern science.

P. O. Johnson[25] suggests that the "newer types of experimental design and the correspondingly appropriate statistical analysis offer new promise for the scientific study of educational problems." He summarizes the research carried out by himself and his students as an illustration of how some problems in the teaching of science can be approached and solved in a scientific manner.

How much longer will the science teachers preach the glory, power, and virtues of the scientific method and display unscientific behavior in the classroom? Has the scientific method become the incurable delusion of the science teachers? The scientific method cannot be taught as a static, isolated concept. The dynamic nature of science demands a dynamic and experimental approach to the teaching of science.

The author wishes to acknowledge his gratitude to Dean J. W. Buchta for helpful criticisms of the manuscript, and to Professors R. Beck and M. Graubard for many helpful suggestions.

[24] Quoted in the Editorial Comments, *J. Higher Educ.*, **18**, 162 (1947).

[25] Johnson, "The scientific study of problems in science education," *Science Educ.* **29**, 175–180 (1945).

Fellowships Administered by the National Research Council

Announcement is made of three types of fellowship administered by the National Research Council.

Joint Fellowships in the Natural and Social Sciences

The Joint Fellowships in the Natural and Social Sciences of the Social Science Research Council and the National Research Council are supported by a grant from the Carnegie Corporation of New York. They are open only to citizens of the United States. Applicants must produce evidence of training equivalent to that represented by the degree of Ph.D., and of unusual talent for research and investigation.

All fields of the natural and social sciences are open to applicants for the Joint Fellowships. Holders will be required to devote their full time to advanced study and research for a period of two years. Six weeks annually will be allowed for vacation. A Fellow may not engage in work for remuneration or receive aid from another appointment, scholarship or similar grant during the tenure of the fellowship.

To receive consideration at the next meeting of the Joint Fellowship Board, applications must be filed on or before February 1, 1949. Applicants must supply a proposed program of study and research, and must indicate the institution in which they desire to study. Awards will be made as soon as possible after March 15, 1949.

The annual stipends, which will be determined in every case by the Joint Fellowship Board, will be in the range $2500 to $5000. Requests for application blanks or for additional information should be addressed to the Joint Fellowship Board in the Natural and Social Sciences, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

Atomic Energy Commission Postdoctoral Research Fellowships in the Physical Sciences

It is the purpose of these fellowships to enable men and women to gain advanced training and research experience in one of the physical sciences: physics, chemistry, geophysics, mathematics, metallurgy, astrophysics, and engineering. The research to be pursued by a Fellow should preferably be concerned with a problem which will permit unrestricted publication of results. The fellowships are supported by the Atomic Energy Commission and the selection of Fellows will be made by the AEC Postdoctoral Fellowship Board in the Physical Sciences of the National Research Council. They are open only to citizens of the United States under thirty-five years of age at the time of appointment. A Fellow must have had training in some branch of the physical sciences equivalent to that represented by the Ph.D. or Sc.D. degree, and must have demonstrated superior ability in research.

Applications to be considered at the Spring (1949) meeting of the Board must be completed by February 15, 1949. The basic stipend is $3000 per year. Larger stipends as determined by the Board may be granted when deemed necessary on the basis of family responsibilities.

Requests for application blanks or for additional information should be addressed to the Fellowship Office, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

(*Continued on page 34*)

The Scientist and Government Research

Eric A. Walker
Ordnance Research Laboratory, The Pennsylvania State College, State College, Pennsylvania

MANY scientists dislike Government service and seem to feel that civil service work carries with it some inherent stigma. The reasons for this feeling are not immediately apparent; and if one reaches for them, they become quite elusive.

Being confronted with this situation some years ago, the author gathered information from a limited number of typical industrial and governmental scientific laboratories, hoping that he could, by analyzing the data, reach some general principles that could be applied by civil service to obtain high grade personnel.

In trying to classify the reasons why many scientists do not care for Government service, one thing is noted very early. That is, although employers often claim that they are unable to get good men because of salary limitations, the question of salary is rarely mentioned by the scientist himself. Very few instances were found in this study where a scientist declined a civil service appointment because the initial pay was too low, the future was uncertain, or the possibility of eventually obtaining a high salary was not great enough. The very lack of complaint on this score leads one to believe that salaries in civil service cannot be far out of line with similar positions in colleges and industry.

To justify the contention, the salaries for a number of colleges, engineering concerns, and Government laboratories are shown in Table I. It is not claimed that this table covers enough cases to be statistically accurate, but it is the feeling of the author that sufficient cases are shown to determine trends. Several salient features of the table should be commented on in more detail.

The system for compiling this table is as follows: The classifications P-1 through P-9 are chosen only as a usable system by which salaries can be differentiated without using actual figures. Some laboratory directors are somewhat reluctant to give salaries in dollars. Each person contributing to this table was asked to designate only the number of people on his staff who fell into the P-1, P-2, or other range. Therefore, the group of people who had salaries between $2644 and $3397 is listed as equivalent to P-1. The distribution used takes no account of the age nor training of the various laboratory members. It shows only the distribution of salaries.

If one uses as a basis for argument all Government scientists listed in column E, it can be seen that the bulk of these men fall in the range of P-2 to P-4. In other words, by far the largest number of Government scientists earn salaries between $3400 and $6000, with a very small percentage reaching $10,000 and only a few isolated individuals listed above that figure.

Compared with these, Navy laboratories, employing approximately 3100 scientists, show a shift in the upward direction. More Navy scientists are included in the P-3 and P-4 groups and fewer in the lower paid P-2 group. If one compares this distribution with a very large industrial corporation including both engineers and research staff as in column C, one finds a very concentrated grouping at about $4000 to $5000. This concentration reduces appreciably the number of men below $4000 as well as those above $5000. One significant difference does not show in the table. In this large industry, and quite probably in similar concerns, there are a few individuals employed for their scientific ability who far exceed the limits of the P-9 classification. They are the important cogs in the industrial scientific machine. This point will be discussed later, since it may be very important in considering the management of scientific laboratories.

Data from two other laboratories which are somewhat similar are given in columns A and B. Column A designates a large laboratory operated by private management and one which derives its funds entirely from Government sources. Laboratory B, on the other hand, is a large research foundation deriving its funds in part from the Federal Government and in part from industry. In each case the management can be

called non-civil service. It will be noted in both these cases that there is a shift towards the upper salary levels in comparison with column E. For instance, in the group from approximately $8200 to $9400, these two laboratories have 6 and 8.6 percent respectively, while in the Government and industrial laboratories there are approximately 2 percent. There is also a small but probably significant number of men who are above the P-9 group.

These differences between the A, B, and C laboratories and the Government service become less startling when one considers the salary range of scientists working for colleges and universities. These are shown in column F. This group includes salaries of teachers and research workers in engineering schools whose salaries are usually at least equal to, if not somewhat higher than, those of similar professors in the liberal arts or classical courses. It is seen that the largest fraction of the staff has salaries of less than $4000 per year. These are normalized to a twelve-month basis. If a faculty member is paid for a lesser period, a correction has been made to reach an equivalent salary. It will be noted that very few college employees receive salaries in excess of $6000 per year.

Since we all admit that a great deal of research, and good research, is done by college and university staffs, it is clear that salary alone is not the reason that scientists are reluctant to enter Government service. It is generally recognized that many intangible compensations in college and university employment make the positions attractive and must be counted among the inducements to serve on a college staff. From this, one concludes that the management and administration of Government laboratories should determine if they cannot offer such inducements.

TABLE I. Distribution of salaries in certain laboratories.

Designation	Salary range* (dollars/yr)	A 1948	B 1948	C 1947	D 1947	E 1944	F 1948
P-1	2644–3397	2.0%	0.0%	0.2%	4.5%	9.1%	26.0%
2	3397–4149	7.5	18.8	5.0	15.1	24.1	36.0
3	4149–4902	19.0	17.8	42.0	34.9	26.7	23.0
4	4902–5905	27.0	29.4	17.0	24.0	19.5	10.0
5	5905–6862	25.0	16.3	12.0	12.0	12.0	3.5
6	7102–8059	12.0	8.6	8.0	6.0	5.4	.5
7	8179–9376	6.0	8.6	2.0	2.2	1.9	1.0
8	9975–10,000	0.6	0.5	1.0	1.16	.7	.5
9	12,500	Small	Small	Small	Small	Almost none	None
Total in group		350	220	1200	3100		250

Key to Table I.

A—A large Government supported privately operated laboratory.
B—A large publicly supported research foundation.
C—A large commercial company's engineering and research staff (as of 1947).
D—Eight Navy civil service laboratories.
E—All Government scientists (as of 1944).
F—Several groups of college-employed scientists.
*—The salary range given for the various civil service ratings is that which was obtained as of January 1, 1948. Since that time, the President has signed a bill which authorizes an increase of $330 in each grade from P-1 through P-7.

Security

In many instances the matter of security is a bothersome one. Some newspaper articles give the impression that scientists, as a whole, are opposed to security measures. This is quite untrue. Scientists, perhaps more than any other group of citizens, are aware of the need for some security measures and systems. A small group of scientists adopted voluntary security measures during the early stages of World War II. At the April, 1940, meeting of the Division of Physical Sciences of the National Research Council, a censorship committee was formed to control publication of material pertinent to the manufacture of the atomic bomb. The committee was organized chiefly to curtail publication of articles on uranium fission but its field of activity expanded later to control publication policy in all fields of possible military interest. Scientists, however, object strongly to rules which treat the scientist as a different sort of person who must be watched more than the average citizen. They object, also, to rules which are patently unenforceable.

The origin of most of the objections lies in the fact that most rules are formulated to apply in the same manner to all individuals and all possible situations. It is perfectly true that there is a rule against homicide, which must apply to every citizen of the country under all circumstances. Few of us will deny that this is a good rule. Quite probably we must have universal regulations prohibiting specific acts which might jeopardize the national security. In some cases the scientist is keenly aware of, and properly or improperly resents, general rules which may work a distinct hardship in individual cases. Furthermore, these regulations are often administered and enforced by men who fail to understand the purpose behind them, or who do not feel able to make justifiable exceptions. For instance, in the interest of national security the Armed Services

may make a rule prohibiting the indiscriminate collection of data concerning the whereabouts of large groups of scientists. A little thought shows that much could be deduced from a collection of facts giving the names of a group of prominent scientists, their training and scientific interests, and their geographical whereabouts at any one time. This is easy to compile. Yet from such data one can determine the scientific problems a country considers important. Therefore, one might be able to plot the scientific progress of a problem similar to the development of the atomic bomb.

However, the enforcement of such a regulation might be entrusted, under limited conditions, to an individual who has no knowledge of the reasons for the rule, nor why and how it should be enforced. This would lead to a situation like the following: The National Research Council, let us suppose, in preparing an inventory of the nation's scientific resources, enlists the assistance of the group which compiles a document like *American Men of Science*. The reason for so doing is obvious. In sending out biographies for an official check, the National Research Council includes a long and somewhat involved questionnaire. The results of this inquiry are to be used for official purposes only, but to be useful they require detailed answers on the part of the scientists. Unfortunately, this questionnaire is seen by a Security Enforcement Officer who immediately decides that it violates the security provisions which the Armed Forces have established. So, he sends a letter, apparently without further consultation with his superiors, which says: ". . . 3. There is another questionnaire being distributed to be used in connection with the eighth edition of 'American Men of Science.' . . . All contractors and their employees are cautioned against including any classified information in their replies. Particular warning is given to the questions 'Item 14, 17, 18, 19, and 23' so that no classified information will be thus revealed. . . . It is earnestly requested that any such questionnaires received by you or your corporation be promptly referred to this office." Thus the Armed Forces successfully prevent themselves from obtaining the data they want.

Such an application of the security rule, although only cautionary in nature, cannot but annoy the scientist, and in many cases, makes him determined to apply his own security laws, rather than depend upon the judgments of others.

It should be pointed out that all scientists who enter Government service pledge themselves not to jeopardize the nation's security. In doing so they nearly always sign the "Espionage Act." When they do this, they, with other citizens, are acutely aware of the pledge they take. They resent enforcement officers assuming that their word is worthless and continually seeking violations, even though the violations are quite unintentional and unimportant.

Such things as the recent Condon inquiry can only make scientists reluctant to accept Government service. Up to the present time no disloyal act has been proved against Condon, and most scientists feel that if they had been in Condon's shoes, they would have done exactly as he did. A public opinion-collecting group asked a number of scientists the following question: "How did the manner in which the charges were made against Condon affect your willingness to accept responsible Government positions?" The answers were as follows:

> "It made me decide to decline any such offer." (12 percent)
> "Made me reluctant to accept." (63 percent)
> "Had no affect." (23 percent)

Notice that it was the ***manner*** in which the charges were made that was questioned, not the investigation itself.

One problem often cited by those who administer the work of Government scientists is the problem of obtaining clearance to publish. It is believed that such difficulties prevent proper recognition, and thus the attraction of Government service is reduced. As a matter of fact, in answer to inquiries, very few scientists have ever given this as a reason for disliking Government service. Perhaps the difficulty is not very real.

In many respects the Armed Forces are extremely liberal in the matter of publication. If a scientific problem can be divorced from classified matters (and usually it can be) then the Armed Forces show little or no reluctance to permit publication. On the other hand, because of the

resources placed at the disposal of scientists by the Government laboratories, very often they are able to publish more material than they could have done in private life.

In answer to a direct question posed to many scientists who had been in Government service, or who contemplated working for the Government, few, if any, have ever said they resented or objected to the necessity for security. However, many have objected to the method in which security matters are enforced.

The Dual-Control Managerial System

In many Government laboratories, especially those controlled by the Armed Services, it seems necessary to have a dual system of control. The organization is set up in terms of a number of organizational boxes designated as desks or code numbers in exactly the same manner as one sets up the organization for a civilian laboratory. However, it appears to be an inviolable rule in our Armed Services that such desks, and the authority that goes with them, can reside only in an officer in uniform. In many cases these positions should be occupied by scientists. So the situation is alleviated by employing a scientist as an aide to the uniformed incumbent. The scientist then finds himself in the position of having to prepare all technical information, reports and letters, but having them signed by his counterpart in whom rests the authority to sign. This is a clear division of responsibility and authority; the authority residing in the uniformed officer, and the responsibility with the scientist.

The system appears to be predicated on the assumption that the scientist, being a queer sort of individual, cannot be trusted with such authority. Such a belief, if existent, is almost insulting. If the job is a scientific one, then the job should be filled by a scientist who has both the authority and the responsibility even though this requires a "break" in the chain of command and necessitates having a scientist in a position where he has authority over the uniformed officer. Such a system has been tried in other countries, and it seems to be quite workable and acceptable. There is no reason why it should not be adopted by our Armed Services. If it is adopted, more scientists would be less reluctant to accept Government service.

Administrative Objections

The majority of the complaints that scientists have against Government service seem to follow no pattern at all. Yet, in many cases the complaints are of such long standing and stand out in such relief in the scientists' minds, that the complainants are willing and, indeed, anxious to talk about the matter. Many of them were asked if they would care to put their complaints in writing—and did so. From these many complaints several are selected to show their caliber:

> One scientist complained that during a trip which he had taken at his employer's insistence he had found it necessary to travel a distance of thirty miles between two small towns. Upon turning in his expense account he found that he should have traveled by train whereas he actually traveled by bus. Because of this he charged the Government twenty-six cents more for this portion of the trip than he should have done. Admittedly, the scientist should have dropped the matter there and accepted the loss of twenty-six cents. Instead, he started a long correspondence to prove the validity of his actions. In this correspondence a great deal more than twenty-six cents was wasted on postage, paper, secretarial time, and the scientist's time in the almost endless interchange of letters. This finally resulted in the scientist's leaving Government service without his twenty-six cents. This seems to be a nonsensical procedure, and one which eventually lost for the Government the services of a competent scientist.

Other examples of opinions, if not reasons, are quoted verbatim from their written statements:

> "A large amount of work done by former scientists now on the Government pay roll is not scientific work. Many of the undertakings are of a coordination, administrative, or management nature. Opportunities of scientific investigation are few and far between, being snowed under by the 'practical' work that prevails under the name of science."

Another writes:

> "Advancement under civil service seems to depend more on how many people you supervise (whether or not they do any useful work), than upon actual ability and achievement."

Still others say:

> "The pyramid system suppresses initiative. Apparently a uniformed officer with a few stripes can get more action toward trying out a crack-pot idea than a scientist with a well-thought-out idea."

"It was almost impossible to spend any time after hours in the laboratories."

"There is a considerable amount of red tape involved in getting anything done. For instance, if I needed a tube not available in our stock room, it would take two days to get it from a supply house ten miles away."

"My general impression of civil service scientists is that they are of relatively low technical quality. They do not work very diligently, and their output is very low. There appears to be little or no incentive for work, and there always appear to be too many men for the work that has to be done."

Lest it appear that only uncomplimentary statements are ever made about civil service, let me quote from a statement by one man who had worked in two Government agencies. One of these apparently was very bad, and the other was very good. Yet, they were operating under the same general rules and regulations.

"With regard to Laboratory X, the story was very different. Laboratory X has some deadheads. However, it has some very capable men who are properly placed in the organizational setup, and the whole *esprit de corps* of this Laboratory is very high. I believe that this, in large part, is due to the administrative setup wherein individuals are encouraged to go into something of scientific interest to them and something which they have run into, and which is still relevant to the general Laboratory problems."

If one tries to find a pattern for all of these complaints, it must be admitted that all of them are petty, and few, if any, are made necessary by the civil service system itself. All seem to have their origin in the administration of the laboratory where the man works; and, as shown by the last quotation, are not to be found in all laboratories.

The only conclusion one can reach is that the *administration and management of some of the civil service laboratories is not all it should be.* Herein lies the crux of the whole problem.

A good many of the laboratory managers have reached their high level by many years of service, and the establishment of the proper amount of seniority. There is not the active competition for the managerial positions which one finds in industry. There is in the latter the insistence that a man must produce and do well in his job, or be fired. Likewise, there is not in civil service the salary incentive that exists in private laboratories, since, until recently, the financial limitation effective in civil service laboratories was $10,000 per annum. This has been remedied to some extent by the addition of the P-9 classification which allows salaries up to $15,000 per annum, which may aid in part the obtaining of a sound, efficient managerial staff in Government laboratories. Such a staff might find it possible to cut through the so-called red tape and set the rules as they are fitted to the needs of the scientists, thus removing these petty annoyances which loom so large in the minds of the individuals to whom they are made to apply.

Fellowships (*Continued from page 29*)

RCA Predoctoral Fellowships in Electronics

It is the purpose of these fellowships to give special graduate training and experience to young men and women who have demonstrated marked ability in the general field of Electronics—either as a branch of electrical engineering or as a part of the general field of physics. The fellowships are supported by the Radio Corporation of America, and the selection of Fellows will be made by the RCA Fellowship Board of the National Research Council.

A Fellow must be a citizen of the United States who has demonstrated ability and aptitude for advanced work, and who has training in Electronics equivalent to that represented by one year beyond the bachelor's degree in a university of recognized merit in this field. He will be expected to work on scientific problems related to Electronics. A Fellow may not engage in work for remuneration or receive aid from another appointment, fellowship, scholarship or similar grant during the tenure of the fellowship.

To receive consideration for tenure during the academic year 1949–50, applications must be filed on or before February 1, 1949. Awards will be announced about March 15, 1949. Requests for application blanks or for additional information should be addressed to the RCA Fellowship Board, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

The Millikan Oil Drop Experiment in Laboratory Courses

O. Oldenberg
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts

THE purpose of this paper is to call attention to a logical difficulty which appears in the determination of the charge of the electron by the Millikan oil drop experiment[1] as it is conventionally carried through in many laboratories. In an advanced laboratory course the student should find out how the fundamental constants of atomic physics are derived from experiments. He should not be allowed unlimited freedom to intersperse among the measurements numerical values of constants taken from a book. In the measurement of the charge of the electron it is apparently fair to take for granted the characteristic constants of air, that is, its *viscosity* which enters into Stokes' law and the *correction factor* which adapts this law to very small droplets. However, the attention of the student must be called to the fact that our knowledge of these two constants of air comes from fundamentally different origins. While the measurement of the viscosity is carried out by an independent experiment (for example, the flow of air through a capillary tube), the correction factor can be determined only by the Millikan oil drop experiment itself. Millikan discovered that *large* droplets yielded consistent values of the charge of the electron, but *small* droplets yielded values apparently increasing with decreasing droplet size. Millikan postulated a constant value for the fundamental charge and attributed the deviation to a failure of Stokes' law; for droplets of a diameter comparable to the mean free path of molecules in air he introduced a correction factor into this law. Using his data on large and small droplets he simultaneously evaluated two unknowns, the charge of the electron and the correction factor of Stokes' law.

In a students' laboratory, the large droplets which do not necessitate a correction cannot be used easily because they require a large plate distance (15 mm) and correspondingly high voltage (up to 6600 volts). Practically, one is restricted to smaller droplets, for which the correction factor is appreciable. But giving the numerical value to the student completes a vicious circle since this value cannot be determined independently. Therefore, the conventional laboratory method of taking the correction factor from a book is as logical as teaching the student how to solve a system of two equations with two unknowns but making it easier by giving him one of the unknowns.

In practice, it is possible to arrange the laboratory experiment so that both unknowns, the charge on the electron and the correction factor, can be satisfactorily determined by skillful students within three or four laboratory periods. Fortunately the correction enters into the equations only as a factor by which the values of the charge on a droplet as derived directly from Stokes' law must be multiplied. Consequently the following procedure is recommended. A droplet is selected whose rate of free fall is most easily measured (falling, for example, 3 mm in 30 or 40 sec). Using this droplet (or several droplets of nearly the same size) and varying the charge several times, many measurements of the charge are made disregarding the correction.

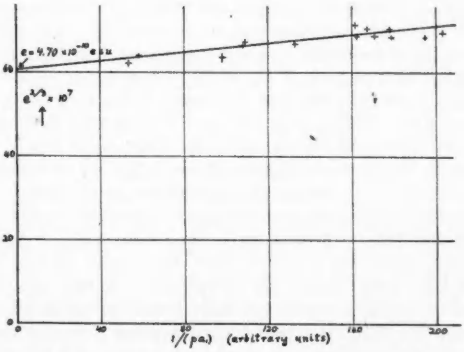

FIG. 1. Correction of Stokes' law in the Millikan oil drop experiment. The zero mark of the ordinate scale is included intentionally in order to demonstrate the relative magnitude of the correction.

[1] R. A. Millikan, *Electrons, protons, photons, neutrons, mesotrons, and cosmic rays* (University of Chicago Press, 1947), p. 99. The simplest performance of the experiment is described by J. B. Hoag and S. A. Korff, *Electron and nuclear physics* (Van Nostrand, 3rd ed.), p. 14, and G. P. Harnwell and J. J. Livingood, *Experimental atomic physics* (McGraw-Hill, 1933), p. 102.

Such a long set will demonstrate the most important fact that the observed charges are approximately multiples of a fundamental value. The charge of the electron shows up as the smallest difference between two observed charges of the droplet. The value thus derived from observations of a few small droplets will be larger than the accepted value of the electronic charge.

Next, in order to determine Millikan's correction to Stokes' law, a few determinations are made with larger and with smaller droplets. These results are then interpreted following Millikan's own procedure. He discovered that the apparent value of the electronic charge e_1 (derived on the basis of Stokes' uncorrected law) is related to the true value e by the equation

$$e_1^{\frac{2}{3}}=e^{\frac{2}{3}}(1+b/pa)$$

where p is the pressure of the gas through which a drop of radius a is falling, and b is an unknown factor. For the determination of the unknown factor b it is sufficiently accurate to apply the values of the radii a as derived from the uncorrected law. This equation suggests that the $\frac{2}{3}$ power of the apparent charge e_1 could usefully be plotted against $1/pa$. Such a graph allows an extrapolation to very large droplets for which $1/pa$ is 0, and the apparent charge e_1 is identical with the true charge e. From the slope of the curve the unknown b is derived. An example is given in Fig. 1.

The experiment as described here has been performed in the Laboratory Course in Atomic Physics, offered at Harvard University, for four or five years. It is true that less skillful students fail to obtain satisfactory values of the correction factor. For such students the simplified experiment, disregarding the correction altogether, is of positive benefit because they obtain from their own experiments evidence for the atomic structure of *electricity*, even though the value of the electronic charge so derived is affected by a systematic error (in practice, up to 15 percent). On the other hand, the corresponding evidence for the atomic structure of *matter* (the laws of constant and multiple proportions) is so much more involved that it would be difficult for students to have the satisfaction of doing all the experiments on which these laws are based.

Experiment *vs.* the Authority of the Ancients

I cannot refrain from marvelling that Sarsi will persist in proving to me, by authorities, that which at any moment I can bring to the test of experiment. We examine witnesses in things which are doubtful, past, and not permanent, but not in those things which are done in our own presence. If discussing a difficult problem were like carrying a weight, since several horses will carry more sacks of corn than one alone will, I would agree that many reasoners avail more than one; but discoursing is like coursing, and not like carrying, and one barb by himself will run farther than a hundred Friesland horses. When Sarsi brings up such a multitude of authors, it does not seem to me that he in the least degree strengthens his own conclusions, but he ennobles the cause of Signor Mario and myself, by showing that we reason better than many men of established reputation. If Sarsi insists that I must believe, on Suidas's credit, that the Babylonians cooked eggs by swiftly whirling them in a sling, I will believe it; but I must say, that the cause of such an effect is very remote from that to which it is attributed, and to find the true cause I shall reason thus. If an effect does not follow with us which followed with others at another time, it is because, in our experiment, something is wanting which was the cause of the former success; and if one thing is wanting to us, that one thing is the true cause. Now we have eggs, and slings, and strong men to whirl them, and yet they will not become cooked; nay, if they were hot at first they more quickly become cold; and since nothing is wanting to us but to be Babylonians, it follows that being Babylonians is the true cause why the eggs become cooked, and not the friction of the air, which is what I wish to prove. Is it possible that in traveling post, Sarsi has never noticed what freshness is occasioned on the face by the continual change of air? And if he has felt it, will he rather trust the relation by others of what was done two thousand years ago at Babylon, than what he can at this moment verify in his own person? I, at least, will not be so wilfully wrong, and so ungrateful to nature and to God, that having been gifted with sense and language I should voluntarily set less value on such great endowments than on the fallacies of a fellow-man, and blindly and blunderingly believe whatever I hear, and barter the freedom of my intellect for slavery to one as liable to error as myself.—Galileo Galilei, *Il Saggiatore*, 1623. Translated by Professor Giorgio de Santillana.

On the Shearing Stress in a Viscous Fluid Across a Surface Normal to the Lines of Flow

G. M. Volkoff and D. S. Carter*
University of British Columbia, Vancouver, British Columbia, Canada

It is well known that in a viscous fluid in a state of steady laminar flow there exists a shearing stress across any surface which separates two adjacent layers of the fluid moving with different velocities. In fact, the emphasis laid in many standard elementary treatments[1] of viscosity on the connection between the relative slipping of adjacent layers and the resultant viscous drag tends to leave the beginning student with the impression that *relative slipping is a necessary condition for a shearing stress between adjacent layers.* Thus the beginner easily visualizes the shearing stress across a cylindrical surface inside, and coaxial with a capillary tube through which laminar flow of a fluid is taking place; but when asked whether any shearing stress exists across an element of surface perpendicular to the axis of the tube, he is likely to reply: "No, because this surface is perpendicular to the lines of flow, and hence no relative slipping of fluid on the two sides of the surface takes place." When, at a later stage, the same student is confronted with the general proof that the stress tensor is necessarily symmetric and that, therefore, there must be a shearing stress across a surface normal to the lines of flow wherever there is a shearing stress across a surface separating two layers of different velocities, he is at a loss to explain this fact in terms of his customary ideas. We shall, therefore, present an elementary physical picture, based on kinetic theory, of the simultaneous existence of both kinds of shearing stress.

The elementary kinetic theory derivation of the coefficient of viscosity of a gas, as presented in most standard texts,[2] is carried out by finding the time rate of transport of momentum across unit area of a surface separating two adjacent gas layers which are "drifting" in the same direction with different speeds; that is, by finding the shearing stress due to "viscous drag." In this usual treatment no asymmetry in the angular distribution of molecular velocities is initially assumed in calculating the number of molecules crossing a given area, and the existing asymmetry is only then taken into account by endowing each molecule with an additional small component of momentum due to its drift velocity, which is assumed to be small compared to the mean speed of thermal motion of the gas. Such a procedure yields a non-zero answer for the shearing stress across a surface *parallel* to the lines of flow, but would give no shearing stress if applied to a surface *normal* to the lines of flow.

In this article we first give a rough qualitative argument to show that it is essential to consider from the outset the existence of the asymmetry in the angular distribution of molecular velocities in order to demonstrate the existence of a shearing stress across a surface normal to the lines of flow. We then formulate our argument quantitatively, and set up a general expression for the rate of transfer of momentum across any unit area in a gas with no limitations on the model used to describe the gas. The symmetry of the stress tensor will follow from this general expression by a purely geometric argument. This is the main point of the present paper. Finally, in order to show that, insofar as the stress across a surface parallel to the lines of flow is concerned, the introduction of the asymmetry in the angular distribution of molecular velocities by the method suggested in this paper leads to the usual answer given by the standard methods, we explicitly evaluate our general expression as applied to a particular frequently used elementary model of a gas (e.g., L. Page, ref. 2). In this model all the molecules are assumed to move with the same speed (equal to the mean thermal speed V of the molecules of an actual gas) on which there is superposed a small drift velocity $\mathbf{u}$. We shall obtain the usual expression $\eta=\frac{1}{3}\rho l V$, where η is the coefficient of viscosity, ρ is the gas density, and l is the mean free path.

---

* Holder of a studentship from the National Research Council of Canada. Now at Princeton University.

[1] G. Joos, *Theoretical physics* (Blackie and Son, 1934), p. 202; L. Page, *Introduction to theoretical physics* (Van Nostrand, 1935), p. 257.

[2] See reference 1, G. Joos, p. 535, and L. Page, p. 343.

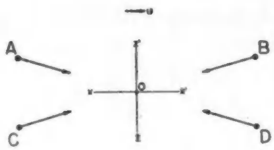

FIG. 1. Contributions of molecular motions to the shearing stresses across xx' and zz'.

For the rough qualitative argument consider the four points A, B, C, D of Fig. 1. They lie in the xz coordinate plane, and are symmetrically situated with respect to the x and z axes. The lines xx' and zz' represent traces on the xz coordinate plane of two small surface elements at the origin O which are perpendicular to the z and x axes, respectively. Consider a gas the molecules of which have a small drift velocity (whose magnitude depends on z only) in the x direction superposed on their random thermal motion. The presence of a drift velocity in the positive x direction means that, of the molecules crossing xx' and zz', fewer arrive from A than from B, and fewer arrive from C than from D. This fact is not of decisive importance in calculating the shearing stress across the surface represented by xx'. This shearing stress is given by the excess of the rate at which net momentum in the x direction is carried across xx' by the molecules crossing xx' from above (from points such as A and B), over that due to the molecules coming from below (from points such as C and D). In calculating the net component of momentum in the x direction crossing xx' due to molecules coming from A and B, it is immaterial whether we directly adopt the usual procedure of saying that *equal* numbers of molecules come from A and from B each carrying the *same average excess* of x-momentum due to the drift velocity, or whether we first actually introduce the asymmetry in the number of molecules coming from points such as A and B, and then wipe out all but the above net effect of the asymmetry by pairing off all points such as A and B and integrating over the upper half-space. The two procedures give the same results because the contributions of the molecules from A and B to the transfer of x-momentum across xx' are *additive* insofar as their drift motion is concerned, and *subtractive* insofar as their thermal motion is concerned. The combined net contribution of molecules from C and D is subtractive with respect to the combined net contribution of molecules from A and B, so that there will be an over-all net transfer of x-momentum across xx' only if there is a gradient in the z direction of the drift velocity in the x direction. When, however, the transfer of z-momentum across zz' is considered, the z-momentum of each molecule is entirely due to the thermal motion of the gas. The contributions of the molecules coming from A and B to the transfer of z-momentum across zz' are subtractive. If no asymmetry in the number of molecules crossing zz' from A and from B were assumed, then the two contributions would exactly cancel. The excess in the number of molecules crossing zz' from A over that from B is the essential feature of the process responsible for the net transfer of *downward* momentum to the right across zz'. Similar considerations applied to points C and D show that net *upward* momentum is being transferred to the right across zz'. The combined contribution of molecules from A and B being opposite to that from C and D, a net transfer of z-momentum across zz' will take place only if there is a gradient along the z-direction of the drift velocity in the x-direction, which ensures that the asymmetry in the numbers of molecules coming from C and D is not so pronounced as that due to molecules from A and B, so that there will be a net transfer of downward momentum to the right across zz'. The above qualitative argument may be restated as follows: While no net transport of molecules takes place across xx', there is a net transport of x-momentum because the average drift momentum is different on the two sides of xx'. In the case of zz' there is no difference in the numbers of molecules having a given z component of momentum due to thermal motion on the two sides of zz' but there is a net transport of such molecules across zz' due to the drift motion, and a resultant net transfer of z-momentum by such molecules. If all positive and negative values of z-momenta are summed there will be an over-all net transfer of z-momentum because the existence of the drift velocity gradient ensures that a greater net transfer of downward momentum will take place across zz' than of upward momentum.

To formulate the above rough arguments quantitatively, choose a point O in the gas which lies on a surface across which the stress is to be calculated. Choose a rectangular coordinate system with origin at O and such that the unit vector $\mathbf{k}$ in the z direction is normal to the surface at O. Then $d\mathbf{S}=\mathbf{k}dS$ is an element of the surface at O. Let P be a second point in the gas and let $\mathbf{r}$ be the vector from O to P. If the gas is in a steady state, the number of molecules which cross $d\mathbf{S}$ in the negative direction per unit time with speeds between v and $v+dv$ after having suffered their last collision in a volume element $d\tau$ situated at P may be expressed in the form

$$F(\mathbf{r}, v)d\tau dv d\Omega, \tag{1}$$

where $d\Omega$ is the solid angle subtended by $d\mathbf{S}$ at P, and $F(\mathbf{r}, v)$ is a function whose form depends upon the particular gas model considered. But if $\mathbf{r}_1$ is the unit vector in the direction of $\mathbf{r}$,

$$d\Omega=\frac{\mathbf{r}_1\cdot\mathbf{k}}{r^2}dS, \tag{2}$$

and hence the number of molecules crossing unit area of our surface in the direction of $-\mathbf{k}$ is

$$F(\mathbf{r}, v)\frac{\mathbf{r}_1\cdot\mathbf{k}}{r^2}d\tau dv. \tag{3}$$

But the velocity of each of these molecules is effectively $-v\mathbf{r}_1$, and therefore the momentum they carry across unit area is

$$-mv\mathbf{r}_1 F(\mathbf{r}, v)\frac{\mathbf{r}_1\cdot\mathbf{k}}{r^2}d\tau dv, \quad (4)$$

where m is the molecular mass.

The total number of molecules crossing unit area in the negative direction, and the total stress exerted by the gas above our surface on that below, may be found by integrating Eqs. (3) and (4), respectively, over all space and speeds. Thus the tangential (shearing) stress in the positive x-direction is

$$P_{xz} = -m\int_\tau d\tau \int_0^\infty dv F(\mathbf{r}, v)\frac{(\mathbf{r}_1\cdot\mathbf{i})(\mathbf{r}_1\cdot\mathbf{k})}{r^2}v. \quad (5)$$

Further, the component P_{zx} in the positive z-direction of the stress exerted at O by the gas to the right of the yz plane on that to the left may clearly be obtained from the right hand side of Eq. (5) simply by interchanging $\mathbf{i}$ and $\mathbf{k}$. But this interchange leaves the expression unchanged, so that

$$P_{zx} = P_{xz}. \quad (6)$$

Finally, in view of the arbitrary way in which both the origin O and the axes of our coordinate system may be chosen, Eq. (6) states directly that the stress tensor is everywhere symmetric.

It is clear that if the function $F(\mathbf{r}, v)$ depends only on the magnitude r, but not on the direction $\mathbf{r}_1$ of $\mathbf{r}$, i.e., if there is no asymmetry at all in the directional distribution of molecular velocities, then there is a net transfer neither of molecules nor of momentum across any surface. The presence of a drift velocity $u\mathbf{i}$ which is constant over the whole gas introduces enough asymmetry to give a net transfer of molecules across a surface element in the yz plane. However, since $F(\mathbf{r}, v)$ remains axially symmetric about $\mathbf{i}$, Eq. (5) still gives $P_{xz} = P_{zx} = 0$, because the average value of $\mathbf{k}\cdot\mathbf{r}_1$ over all those values of $\mathbf{r}_1$ for which $\mathbf{i}\cdot\mathbf{r}_1$ (and consequently $F(\mathbf{r}, v)$) remains constant, is zero.

The presence of a gradient of u in the z-direction removes the axial symmetry of $F(\mathbf{r}, v)$ about $\mathbf{i}$, and is responsible for non-vanishing shearing stresses $P_{xz} = P_{zx}$, even though P_{xz} is exerted normally to the lines of flow.

We finally specialize the above results to the simple model of a gas in which there is a constant number N of molecules per unit volume, and such that if the drift velocity were zero, all the molecules would move with the same speed V and with a random distribution over all directions. In assuming a drift velocity which varies throughout the gas, we mean that an observer moving with the drift velocity of a certain point would see those molecules which have just collided near that point moving as though they were in a stationary gas. Thus we assume a drift velocity $u\mathbf{i}$ where

$$u = u_0 + cz \quad (7)$$

such that c is a constant gradient, and $u \ll V$ for all values of z of the order of the mean free path l.

Let r, θ, and ϕ be the polar coordinates of a point P in terms of a coordinate system with origin at O and polar axis in the direction of $\mathbf{i}$. Let v' and v'' be the speeds of a molecule moving away from P as seen by a stationary observer and by an observer moving with the drift velocity $\mathbf{u}$ of the gas at P, respectively; and let the directions of motion of the molecule in terms of polar coordinates with origin at P be denoted by θ' and ϕ' for the stationary observer, and by θ'' and ϕ'' for the moving observer (see Fig. 2). Then if $u \ll v'$, we find that to the first order in u/v':

$$v'' \approx v'(1 - \lambda' u/v'), \quad (8a)$$

$$\lambda'' \approx \lambda' + (\lambda'^2 - 1)u/v', \quad (8b)$$

$$\phi'' = \phi', \quad (8c)$$

where

$$\lambda' = \cos\theta' \quad \text{and} \quad \lambda'' = \cos\theta''. \quad (8d)$$

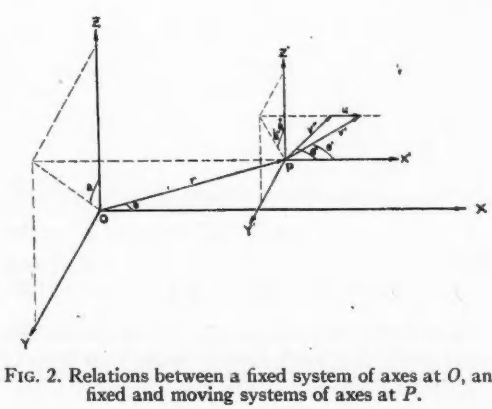

FIG. 2. Relations between a fixed system of axes at O, and fixed and moving systems of axes at P.

Further, if r' and r'' are the distances that the molecule travels during a certain time, as seen by the stationary and the moving observer, respectively, we have

$$r''=r'v''/v'\approx r'(1-\lambda' u/v'). \tag{9}$$

In accordance with approximations made in the elementary treatments, the number of molecules, as seen by the drifting observer, which collide per unit volume per unit time at P, leave within the solid angle $d\Omega''=d\lambda''d\phi''$ with a speed between v'' and $v''+dv''$, and travel a distance r'' without colliding again is taken to be

$$\frac{NV}{4\pi l}e^{-r''/l}\delta(v''-V)d\lambda''d\phi''dv'', \tag{10}$$

where $\delta(v''-V)$ is the Dirac delta-function. But this is equal to the corresponding number as seen by the stationary observer, where the distance is r', the speed is v', and the solid angle is $d\Omega'=d\lambda'd\phi'$ such that

$$d\lambda''d\phi''dv''=J\left(\frac{\lambda'',\phi'',v''}{\lambda',\phi',v'}\right)d\lambda'd\phi'dv'. \tag{11}$$

Evaluating the Jacobian by means of Eqs. (8), we find that to the first order in u/v', the number as seen by the stationary observer is

$$\frac{NV}{4\pi l}e^{-r''/l}\delta(v''-V)(1+2\lambda' u/v')d\Omega'dv'. \tag{12}$$

For molecules moving toward O, we set

$$\lambda'=-\lambda, \tag{13a}$$

$$v'=v \tag{13b}$$

and

$$d\Omega'=d\Omega, \tag{13c}$$

where $\lambda=\cos\theta$ (see Fig. 2). Thus we obtain

$$F(\mathbf{r},v)=\frac{NV}{4\pi l}e^{-r''/l}(1-2\lambda u/v)\delta(v''-V), \tag{14}$$

where, in accordance with Eqs. (8), (9) and (13),

$$v\approx v''-\lambda u \tag{15a}$$

and

$$r\approx r''(1-\lambda u/v''). \tag{15b}$$

In order to check Eq. (14), we now calculate the number ψ of molecules crossing unit area of the xy plane at O per unit time in the direction of $\mathbf{i}$, by substituting $\mathbf{r}_1\cdot(-\mathbf{i})=-\lambda$ for $\mathbf{r}_1\cdot\mathbf{k}$ and $d\tau=r^2drd\lambda d\phi$ in Eq. (3) and integrating. Since it is convenient to integrate over r'' and v'', we also make the substitution

$$drd\lambda d\phi dv=J\left(\frac{r,\lambda,\phi,v}{r'',\lambda,\phi,v''}\right)dr''d\lambda d\phi dv''. \tag{16}$$

Evaluating the Jacobian to order u/v with the help of Eqs. (7) and (15) and the fact that $cz/v\approx cr''\sin\theta\cos\phi/v''$, we find that

$$\begin{aligned}\psi\approx&\frac{NV}{4\pi l}\int_0^\infty dr''\int_{-1}^{+1}d\lambda\int_0^{2\pi}d\phi\\&\times\int_0^\infty dv''e^{-r''/l}\delta(v''-V)\\&\times(-\lambda)(1-3\lambda u_0/v''-4\lambda cr''\sin\theta\cos\phi/v'')\\=&Nu_0,\end{aligned} \tag{17}$$

which clearly is the correct expression.

In order to calculate the shearing stresses $P_{zx}=P_{xz}$, we substitute Eq. (14) into Eq. (5) and use Eqs. (15a) and (16) to obtain

$$\begin{aligned}P_{xz}=P_{zx}\approx&-\frac{NmV}{4\pi l}\int_0^\infty dr''\int_{-1}^{+1}d\lambda\int_0^{2\pi}d\phi\\&\times\int_0^\infty dv''e^{-r''/l}v''\delta(v''-V)\lambda\sin\theta\cos\phi\\&\times(1-4\lambda u_0/v''-5\lambda cr''\sin\theta\cos\phi/v'')\\=&\tfrac{1}{3}NmVlc\\=&\tfrac{1}{3}\rho Vlc.\end{aligned} \tag{18}$$

Thus, in accordance with the general definition of the coefficient of viscosity η:

$$P_{xz}=P_{zx}=\eta(\partial u_z/\partial x+\partial u_x/\partial z), \tag{19}$$

and the fact that for our case $u_z=0$ and $\partial u_x/\partial z=c$, we obtain the customary expression $\eta=\frac{1}{3}\rho Vl$.

We wish to thank Professor F. J. Belinfante for several illuminating discussions of this problem. The junior author (D.S.C.) wishes to thank the National Research Council of Canada for the award of a studentship during tenure of which the above work was done.

On the Principle of Carathéodory

H. A. Buchdahl
University of Tasmania, Hobart, Tasmania

I

In most treatments of thermodynamics the Second Law is stated in one or other of the original forms resulting from Clausius and Thomson. However, Carathéodory, in his axiomatic development of thermodynamics, replaced the traditional statements of the Second Law by what has become known as the *Principle of Carathèodory*. A more widespread knowledge of the methods of Carathéodory[1] seems desirable, not least because of their great didactic value. Experience shows that they can be understood by undergraduates in the second or third year of a course in physics. These methods involve, however, one mathematical theorem (the *Theorem of Carathéodory*), the usual proofs[2] of which are often so unpalatable to the physicist, that the theorem itself may form a serious obstacle to a proper understanding of the whole treatment. We shall therefore give in a subsequent paper an alternative proof of this theorem for the case of three variables, which may be found more attractive than the proof usually offered in text books. It is desirable in the development of any physical theory that there should be a clear-cut division between empirical content and mathematical method. In the case we are considering, once Carthéodory's Theorem is understood as a theorem in pure mathematics, the existence of a certain single-valued function of the variables of state is at once seen to be an immediate consequence of the generalized empirical knowledge which is contained in the Second Law (in the form of Carathéodory's Principle). In the usual treatments, the existence of this function is generally proved with the aid of abstract engines and cycles—a method which may leave some students without much appreciation of what has been proved, and without too clear an understanding of the phenomenological meaning of entropy.

[1] Carathéodory, *Math. Annalen.* **67**, 355 (1909) and *Sitz. d. Preu. Akad. d. Wiss.*, p. 39 (1925).

[2] Born, *Physikalische Ztschr.* **22**, 251 (1921) and *Chandrasekhar, Stellar Structure* (Univ. Chicago Press, 1939), Ch. I, p. 21 and *Carathéodory*, (1909), *loc. cit.*, p. 369.

II. The Principle of Carathéodory

Keeping the preceding remarks in mind, and in order to emphasize the close analogy which exists between the 'physical argument' and the 'mathematical argument' of this treatment we propose to deal in this paper with some general considerations concerning the Principle of Carathéodory. The latter may be stated as follows:

In the neighborhood of any arbitrary initial state J_0 of a physical system there exist neighboring states J which are not accessible from J_0 along adiabatic paths.

This principle thus takes as a starting point the empirical recognition that if two states, J_0 and J, of a given adiabatically enclosed[3] thermodynamic system be prescribed, and granted (i) that the transition from J_0 to J is mechanically possible, and (ii) that such a transition would not violate the demands which the First Law of Thermodynamics already imposes upon it; then the transition from J_0 to J may nevertheless be impossible, while at the same time the reverse transition is possible. We then say that the *thermodynamic weight* of J_0 exceeds that of J.

Let us consider an elementary example. If $J_0 \rightarrow J$ stands for the phrase 'the transition of the system from the state J_0 to the state J,' let $J_0(h_0, t_0)$ be the state of the Joule paddle-wheel apparatus, the contents of the calorimeter being at temperature t_0, and the mass m at height h_0. Let $J(h, t)$ be a second state of the system, where $t < t_0$ and $h > h_0$, in such a way that the energy difference of the contents of the calorimeter, to which corresponds the temperature difference $t_0 - t$, is just accounted for by the potential energy difference of the mass $(=mg(h-h_0))$ in the two states of the system respectively. Then we know empirically that $J_0(h_0, t_0) \rightarrow J(h, t)$ is impossible, notwithstanding the fact that neither the First Law, nor the laws of mechanics would

[3] A system is said to be adiabatically enclosed if a state of equilibrium of the system can be disturbed only by mechanical means, as by shaking, stirring, or the passage of electric currents.

be violated in this transition; $J(h, t) \rightarrow J_0(h_0, t_0)$ is, however, possible.

III. Properties of Systems and Single-Valued Functions

In order to simplify the following considerations, which are of a fairly general kind, let a system K consist of a gas within an envelope,[4] the volume occupied by the gas being v, at a pressure p. We regard the mechanical variables p and v as the independent *variables of state*, i.e., the quantities p and v define the state of the system, and are within certain limits variable at will. In a manner which we need not consider here, the conditions for thermal equilibrium lead us to associate with given values of p and v a number t, such that two such systems K and K' can be in thermal equilibrium if, and only if, the corresponding numbers t and t' are equal.[5] That is, empirical knowledge concerning the thermal equilibrium of physical systems leads to the definition of a single-valued function $t(p, v)$ of the variables of state, which expresses a new property of the system, *viz.*, the property of being, or of not being in thermal equilibrium with another system when the two are brought into non-adiabatic contact. Any other such definitive property,[6] often expressed in the form of a natural law, may similarly be expected to lead to the attachment of a certain number to every given state, i.e., to the definition of a new single-valued function of the variables of state associated with the system, which expresses this property. The First Law of Thermodynamics is an excellent example; it generalizes the result of a very great number of experiments in the statement that *the mechanical work W done by a system in any adiabatic transition between two given states depends upon these states alone, not on the manner of transition.* The definition of a new single-valued function of the variables of state, the energy U of the system, is an immediate consequence of this statement. The term 'quantity of heat' (Q) then appears merely as an abbreviation for the difference between the actual work done in a given non-adiabatic transition and the change in the value of the energy function which occurs in it. Thus if U_0, U are the values of the energy in the initial and final states respectively, then

$$Q = (U - U_0) + W. \qquad (1)$$

IV. The Existence of the Entropy Function

After the preliminary observations of the preceding section we return to the consideration of Carathéodory's Principle. As we have seen, the latter expresses a definitive property of the system, *viz.*, the property that, when adiabatically isolated, the possibility or impossibility of $J_0 \rightarrow J$ depends upon J_0 and J alone, subject to certain other well-defined conditions being already satisfied. Accordingly we may expect the principle to lead to a new single-valued function of the variables of state S,[7] such that S is a measure of the thermodynamic weight of the state J. We call S the entropy of the system. It follows at once, the sign of S being suitably chosen, that $J_0 \rightarrow J$ is possible if $S \geqslant S_0$, and impossible if $S < S_0$; for the condition of accessibility cannot be expressed in any essentially different way in terms of a pair of numbers which must enter into the relations quasi-symmetrically. Moreover, let $S > S_0$; then $J_0 \rightarrow J$ is possible. But having effected $J_0 \rightarrow J$, $J \rightarrow J_0$ is now impossible, for now $S(\textit{final}) < S(\textit{initial})$. That is, $J_0 \rightarrow J$ is irreversible. Clearly $J_0 \rightarrow J$ is reversible only if $S_0 = S$. The last results may be summed up as follows:

A transition of an adiabatically enclosed system is impossible, possible reversibly, or possible irreversibly according as the entropy of the initial state is greater than, equal to, or less than that of the final state.

This at once gives rise to the corollary that in any adiabatic transition of a system the entropy can never decrease. This is the so-called Principle of Increase of Entropy, which shows that unlike mass, energy, charge etc. entropy obeys a one-sided conservation law.

We shall not pursue the physical consequences

[4] The envelope is not to be regarded as forming a part of the system.

[5] The apparent existence of more than one value of t for given p and v, e.g., water near 4°C, would show the incorrectness of the assumption that p, v constituted a sufficient set of independent variables of state.

[6] A property of this type evidently cannot depend on the previous history of the system.

[7] It is conceivable that it could define more than one new function determining the mutual accessibility of different states, but it is difficult to see how this could come about on the basis of a law of the type under consideration. However, it appears that ultimately we must rely upon the confirmation obtained from a mathematical treatment of the problem.

of the Principle of Carathéodory beyond this point; for the elucidation of the phenomenological meaning of entropy as 'transition potential' has been dealt with at sufficient length for our purpose.

V. Plausibility Arguments Based on Carathéodory's Principle

Finally we examine briefly how the considerations above indicate to us how to begin the mathematical formulation of the consequences of Carathéodory's Principle. To do this it is sufficient to consider a system L with three independent variables of state (such as the two aforementioned systems K and K' in thermal equilibrium), which we take to be v, v' and the common temperature t. Now Carathéodory's Principle speaks of arbitrary adiabatic transitions. It applies therefore *a fortiori* to quasi-static[8] adiabatic transitions. During an infinitesimal part of it the work done by L is $pdv+p'dv'$; and since the transition is adiabatic this work must, in virtue of the definition of energy, be equal to the change dU in the energy $U(v, v', t)$ of L, i.e.,

$$(\partial U/\partial v+p)dv+(\partial U/\partial v'+p')dv'+(\partial U/\partial t)dt=0. \quad (2)$$

Thus the quasi-static adiabatic transitions of L are subject to a condition of the form

$$dQ\equiv P(x, y, z)dx+Q(x, y, z)dy+R(x, y, z)dz=0, \quad (3)$$

where P, Q, R are certain functions[9] of the independent variables x, y, z. Interpreting the latter as right angled Cartesian coordinates of a 'picture space' A, every state of equilibrium of L is represented by a point in A. By definition a quasi-static transition must therefore be represented by a continuous curve C in A. If the transition is also adiabatic, C is restricted by Eq. (3). In other words: The 'quasi-static adiabatics' of L are the solution curves of the differential Eq. (3).

But in Sec. IV we tentatively accepted the existence of a certain function S, and we saw that it remains constant in a quasi-static adiabatic transition. That is, as a consequence of the Second Law there exists a function S such that the equation

$$Pdx+Qdy+Rdz=0 \quad (4)$$

implies

$$dS=0, \quad (5)$$

and which has the properties described in Sec. IV. Hence if we are on the right track we may expect that there exists another function $\omega(x, y, z)$, such that[10]

$$Pdx+Qdy+Rdz\equiv\omega dS. \quad (6)$$

We have arrived at the tentative Eqs. (5) and (6) by means of a direct physical 'plausibility argument' based on Carathéodory's Principle. These equations may now be put on a rigorous basis through an application of the Theorem of Carathéodory, the statement, and a new proof of which will form the substance of another paper,[11] in accordance with our original intention of delimiting the mathematical core of the consequences of the Second Law.

[8] A tansition of a system L is said to be quasi-static if, in the course of it, L passes through a continuous series of states of equilibrium. This is equivalent to a reversible transition, which necessarily proceeds at an infinitesimal rate.

[9] Equation (3) preserves its form under any substitution of independent variables. Thus, if the set x, y, z be given as functions of the new set x', y', z', then the equation becomes $P'dx'+Q'dy'+R'dz'=0$, where $P'(x', y', z')=P\partial x/\partial x'+Q\partial y/\partial x'+R\partial z/\partial x'$, etc. Note that the Q here has nothing to do with the symbol for quantity of heat.

[10] For quite unrestricted P, Q, R this is, in general, impossible; in fact, the 'condition of integrability,' $P(\partial Q/\partial z-\partial R/\partial y)+Q(\partial R/\partial x-\partial P/\partial z)+R(\partial P/\partial y-\partial Q/\partial x)=0$, must be satisfied. See Forsyth, *Differential equations* (Macmillan, 3rd ed. 1903), pp. 282–284.

[11] Buchdahl, see article in this issue.

I should like to mention one case where the artist—quite unknowingly—has been able to provide valuable data for science. The clay from which the Greek potter made his beautiful vases more than 2000 years ago always contained some magnetic oxide of iron. At a certain stage of the cooling, after firing, the iron particles are very susceptible to the action of magnetic forces, and orient themselves in the direction of the earth's magnetic field. The direction of this magnetization was fixed permanently when the vase cooled and since we know the vase must have been always in a vertical position during the firing, the scientific man can find the direction of this magnetization and thus fix the inclination or 'dip' of the earth's magnetic field at the time and for the place where the vase was made. By this curious observation we have been able to extend our knowledge of the secular variations in the earth's magnetic field to a remote epoch more than 2000 years before the importance of such measurements was recognized.—E. Rutherford (1932).

On the Theorem of Carathéodory

H. A. BUCHDAHL
University of Tasmania, Hobart, Tasmania

I

THE Second Law of Thermodynamics in the form of the *Principle of Carathéodory* states that if we consider different states of a given physical system then in the neighborhood of any arbitrary state J_0 there are states J which are not accessible from J_0 along adiabatic paths. This principle was considered by the author in another paper[1] from a physical standpoint. References to it will be denoted by the symbol P. As announced there, we now propose to give a straightforward analytical proof of the Theorem of Carathéodory.[2] This is most easily stated in terms of a picture space A, with rectangular coordinates x, y, z, of which a certain finite region D is contemplated. The theorem then takes the form:

In the neighborhood of any arbitrary point G_0 there are points G which are not accessible from G_0 along solution curves of the equation

$$P(x,y,z)dx+Q(x,y,z)dy+R(x,y,z)dz=0, \quad (1)$$

if, and only if, the equation is integrable.

The equation is called integrable if it is equivalent to a single finite algebraic relation

$$F(x,y,z)=\text{const.}, \quad (2)$$

that is, if there exist functions $\lambda(x,y,z)$, $F(x,y,z)$ such that[3]

$$Pdx+Qdy+Rdz\equiv\lambda dF. \quad (3)$$

The theorem, whose close verbal resemblance to the principle should be noted, allows of an immediate rigourous mathematical formulation of the consequences of the Second Law. For in terms of the picture space A the quasi-static adiabatics of a system L with three independent variables of state are the solution curves of a differential equation of just the form of Eq. (1), (cf. P, Sec. V). And if we consider *quasi-static* transitions of L the stated principle asserts in particular: in the neighborhood of any arbitrary point G_0 there are points G which are not accessible from G_0 along solution curves of Eq. (1). The theorem then immediately provides us with Eq. (3): and this has already the form of Eq. P (7). We return to these considerations briefly in Sec. IV of this paper.

II. Change of Variables

It is not difficult to see that the solution of Eq. (1) is a set of curves; for the equation merely prescribes that at any point G a certain line element with components dx, dy, dz shall be perpendicular to the given vector with components P, Q, R at G. Hence any curve which is such that the tangent to it at any point is perpendicular to the vector (P, Q, R) at that point is a solution curve of Eq. (1).[4]

To proceed with the proof of the theorem it is convenient to carry out a certain change of variables, for which purpose we determine a pair of functions $u(x,y,z)$, $\mu(x,y,z)$, such that

$$\left.\begin{aligned} P&=(1/\mu)(\partial u/\partial x),\\ Q&=(1/\mu)(\partial u/\partial y).\end{aligned}\right\} \quad (4)$$

Such functions can always be found. Thus $\mathbf{u}$ is a solution of the differential equation

$$Q(\partial u/\partial x)-P(\partial u/\partial y)=0, \quad (5)$$

in which z is regarded as a constant. Writing

$$\mu R-(\partial u/\partial z)=Z, \quad (6)$$

Eq. (1) takes the form

$$du+Zdz=0. \quad (7)$$

It is assumed that in the range D the quantities P, Q, R and μ are such that u and Z are single-valued, finite and continuous functions of x, y, z and possess finite and continuous first partial

[1] Buchdahl, see article in this issue.
[2] We restrict the theorem to the case of an equation in three variables.
[3] If $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$ are quite unrestricted such functions do not in general exist. In fact, they exist only if the 'condition of integrability,' *viz.*, $\mathbf{P}(\partial\mathbf{Q}/\partial\mathbf{z}-\partial\mathbf{R}/\partial\mathbf{y})+\mathbf{Q}(\partial\mathbf{R}/\partial\mathbf{x}-\partial\mathbf{P}/\partial\mathbf{z})+\mathbf{R}(\partial\mathbf{P}/\partial\mathbf{y}-\partial\mathbf{Q}/\partial\mathbf{x})=0$, is satisfied identically.
[4] See also the excellent discussion of A. R. Forsyth, *Differential equations* (MacMillan, 6th ed., 1943), pp. 318–324.

differential coefficients with respect to x, y, z. In place of x, y, z we now adopt u, y, z as independent variables, (physically: change of variables of state!); hence

$$Z=Z(u, y, z). \tag{8}$$

Corresponding to this change of variables we now use a picture space A' with rectangular coordinates u, y, z. In accordance with the assumptions just made there will be a reversible one-to-one correspondence between the points in the range D of A and the points in the corresponding range D' of A', and moreover it suffices to prove the theorem for the case of Eq. (7).

III. Proof of Carathéodory's Theorem

If $H_0(u_0, y_0, z_0)$ is the point in D' which corresponds to an arbitrary point G_0 in D, consider how the passage along a solution curve of Eq. (7) from H_0 to a neighboring point H may actually be effected.

(i) First, pass in the plane $u=u_0$ from H_0 to the point H_1. Since, by Eq. (7), $z=const.=z_0$, the coordinates of H_1 are

$$(u_0, y_1, z_0), \tag{9}$$

where y_1 may be chosen at will within D', $(-\sigma \leqslant (y_1-y_0) \leqslant \sigma'$, say; $\sigma, \sigma'>0)$.

(ii) Next, pass in the plane $y=y_1$ from H_1 to the point H_2. Equation (7) now reads

$$du+Z(u, y_1, z)dz=0, \tag{10}$$

the solution of which may be written

$$u=q(z, y_1), \tag{11}$$

where the constant of integration is so chosen that

$$q(z_0, y_1)=u_0. \tag{12}$$

Hence the coordinates of H_2 are

$$q(z_2, y_1), y_1, z_2, \tag{13}$$

where z_2 may be chosen at will within D'.

(iii) Finally, pass in the plane $z=z_2$ from H_2 to the point H, which will have the coordinates

$$q(z_2, y_1), y_3, z_2, \tag{14}$$

where y_3 may be chosen at will within D'.

Now $q(z, y_1)$ is a finite and continuous function of z, y_1; moreover, $\partial q/\partial y_1$ is finite and continuous in D'.[5] Consider the equation

$$\partial q/\partial y_1=0. \tag{15}$$

Two possibilities arise: either Eq. (15) is satisfied identically everywhere in D' or it is not. In the latter case, remembering that Eq. (7), and, therefore Eq. (1), are then not integrable, we choose H_0 such that

$$(\partial q/\partial y_1)_{y_1=y_0}\neq 0. \tag{16}$$

Then—keeping the continuity conditions in mind—there exist positive numbers ϵ_1, ϵ_2 such that for any z_2, $(z_0-\epsilon_1\leq z_2\leq z_0+\epsilon_1)$, we can determine y_1, $(y_0-\sigma\leq y_1\leq y_0+\sigma')$, so that $q(z_2, y_1)$ takes on any prescribed value lying between the limits $u_0-\epsilon_2$, $u_0+\epsilon_2$. Hence if, for some positive number ϵ_3, we take

$$y_0-\epsilon_3\leq y_3\leq y_0+\epsilon_3,$$

we see at once that certainly all points H lying in the range

$$\left.\begin{aligned}-\epsilon_2\leq u-u_0\leq\epsilon_2\\ -\epsilon_3\leq y-y_0\leq\epsilon_3\\ -\epsilon_1\leq z-z_0\leq\epsilon_1\end{aligned}\right\} \tag{17}$$

are accessible from H_0 along solution curves of Eq. (7), in the manner specified in Sec. III, while ϵ_1, ϵ_2, ϵ_3 must, of course, be such that the various points considered above lie in D'.

From the last result it follows immediately that under the conditions stated there are in general inaccessible points in the neighborhood of any arbitrarily chosen initial point only if Eq. (15) is satisfied identically in D'. In that case q and therefore Z are independent of y_1, that is, of y; and Eq. (7) is obviously integrable. Consequently, Eq. (1) is integrable; so that, by Eq. (2), all solution curves passing through G_0 lie in the surface

$$F(x, y, z)=F(x_0, y_0, z_0), \tag{18}$$

for at any point a small displacement must, as we have seen, take place perpendicularly to P, Q, R, i.e., perpendicularly to the normal to the surface at the point. Accordingly, all points G in the neighborhood of G_0 which do not lie on the integral surface of Eq. (18) are inaccessible from G_0 along solution curves of Eq. (1); and hence the theorem is proved.

IV. Absolute Temperature and Entropy

We have now established firmly that for quasistatic transitions of the system L

$$dQ=\lambda dF, \tag{19}$$

where λ and F are functions of the variables of state, [cf. Eqs. (3) and P(4)]. The paper of

[5] Kamke, *Differentialgleichungen* (Becker, Erler; Leipzig 1943, 2nd ed.), p. 35.

Born[6] may be consulted for a simple demonstration that Eq. (19) may be written in the usual form

$$dQ = TdS, \tag{20}$$

where $T(t)$ is a universal function[7] of the thermometric temperature t of L:

$$\frac{d \log T(t)}{dt} = \frac{\partial \log\lambda}{\partial t}, \tag{21}$$

while

$$S = \int \Phi(F)dF, \tag{22}$$

where $\Phi(F)$ is some function of F. In the same paper it is shown that S then indeed possesses all the properties which we expect it to have on the basis of the discussion of P, Sec. IV; and the question of the practical determination of the various functions introduced is briefly considered.

[6] Born, *Physikalische Ztschr.* **22**, 282–286 (1921).

[7] Universal in the sense that, if L' is another system in equilibrium with L, then $\partial \log\lambda'/\partial t = \partial \log\lambda/\partial t$, where λ' is the 'integrating denominator' of dQ', primed and unprimed quantities referring to L' and L respectively. (Note that t is the same for both systems.)

AMERICAN JOURNAL OF PHYSICS VOLUME 17, NUMBER 1 JANUARY, 1949

NOTES AND DISCUSSION

A Simulated Electric Line

REGINALD T. HARLING
St. Lawrence University, Canton, New York

A SIMULATED line is made at very little cost from a few resistors and condensers. It seems to give added interest to measurements of resistance and capacity in an elementary or intermediate electricity laboratory.

Three resistors R_1, R_2, and R_3, are connected in series as shown in Fig. 1 to form one side AB of the line; and three others, R_4, R_5, and R_6, form the side CD. Binding post G represents a ground connection. Condensers C_1, C_2, C_3, and C_4, represent the capacitance of the line. The resistors and condensers may have any values chosen so that the resistances and capacitances are easily measurable with simple apparatus. Switches numbered 1 to 10 are connected as shown.

The apparatus is first given to the student with switches 1, 2, 3, and 4 closed and the others open, representing the line in good condition. The student is told to measure (i) the resistance between A and C with BD shorted and (ii) the capacity between A and C.

By manipulating the switches, the instructor then produces the effect of either (a) a complete break in the line, at any one of four places, (b) a grounded line at any one of four places, or (c) a "cross" in the line at either of two places. The student is asked first to find which of these three types of fault is present. For this purpose he uses either an ohmmeter, or a dry cell in series with a voltmeter or a flash lamp. He is then told to assume that the line is fifty miles long from AC to BD, and by making resistance or capacitance measurements, he finds the distance of the fault from one end.

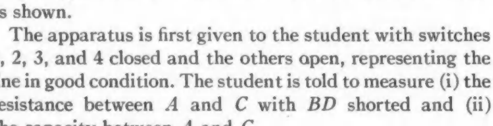

FIG. 1. A simulated electrical line.

It has been found that the experiment arouses much more interest than a similar one in which the object is merely the measurement of a resistance or a capacitance for no apparent reason other than the practice involved. In particular, considerable useful discussion arises concerning methods of making measurements which are within the scope of the equipment on the bench in its present form, but are not possible in the practical case in which only one end of the line is accessible.

Letters to the Editor

THE Editor invites correspondence concerning the subject matter of articles published in the *American Journal of Physics* and concerning other topics of current interest in the teaching of physics. Prompt publication in the Notes and Discussion section of a limited number of communications is assured, but no proof can be sent to authors. Letters should not exceed 500 words in length. They should be addressed to The Editor, American Journal of Physics, Michigan State College, East Lansing, Michigan.

RECENT MEETINGS

New England Section, American Physical Society

The regular fall meeting of the New England Section of the American Physical Society was held at Bowdoin College, Brunswick, Maine, on October 23, 1948. Fifty-two members of the section registered. K. T. Bainbridge, Chairman of the Section, addressed the meeting upon the subject "Isotopic Weights." The program included four contributed papers and a Symposium on Low Temperature Physics. Luncheon was served to members and guests at the Moulton Union. A brief business meeting was held at the close of the morning session. At the business meeting, the following officers were elected for 1949: *Chairman*, MILDRED ALLEN, *Mount Holyoke College; Vice-Chairman*, R. D. EVANS, *Massachusetts Institute of Technology; Secretary-Treasurer*, G. F. HULL, JR., *Dartmouth College; Program Committee*, J. C. STREET, *Harvard University*, and T. SOLLER, *Amherst College*. The program was as follows.

Isotopic weights. K. T. BAINBRIDGE, *Harvard University.*

Physical science for students of the liberal arts. J. J. G. MCCUE, *Smith College.*

Precise measurements of the molar refraction of carbon dioxide and its variation with density. VICTOR H. COFFIN and CLARENCE E. BENNETT, *University of Maine.*

A method of measurement of small periodic displacements and its application to determining the piezoelectric constants of potassium dihydrogen arsenate. THADDEUS NIEMIEC, *Wesleyan University.*

Fringing field corrections for magnetic sector lenses and prisms. K. T. BAINBRIDGE, *Harvard University.*

Symposium on Low Temperature Physics

Fluid properties of liquid helium. A. D. MISENER, *University of Toronto.*

The helium three isotope at liquid helium temperature. H. A. FAIRBANK, *Yale University.*

The surface resistance of superconductors at microwave frequencies. W. M. FAIRBANK, *Amherst College.*

Abstracts of the ten-minute contributed papers will appear in the *Physical Review*.

GORDON F. HULL, JR.
Secretary-Treasurer

DIGEST OF PERIODICAL LITERATURE

Attitude and Education

The prime purpose of education is to inculcate healthy attitudes. This cannot be done any better by the humanist than by the scientist. Indeed, the choice of subjects has little to do with it. Science does not need any more, or less, humanizing than does, say, archeology or etymology. In any case, a right attitude toward life must be acquired, and this can be accomplished mainly through our teachers. Several experiences as a student are cited by the author as evidence that science (and the humanities) can teach more than just facts. One example is that of the college algebra teacher who spent the entire first meeting of the course in writing an infinite series on and on across the blackboard. After commenting that the series could be expanded to the outermost reaches of interstellar space, he said as the bell rang, "Come back Thursday and I'll tell you the secret of adding all these terms." The concept of entropy, for example, is one over which the teacher can romanticize. It can be presented, as can other aspects of science, in a manner which promotes an attitude of humility and awe. We must recognize not only the scholarship of our teachers but their inspirational teaching as well. They must acquire honor and respect as they make detail subservient to principle, and as they inspire in their pupils attitudes of humility, of good will and of readiness to observe beauty wherever it may be found. W. B. WIEGAND, *J. Chem. Educ.* **25**, 362–364 (1948).—R. T. L.

The Glycerol Vapor Vacuum Pump

The efficiency of a vapor pump in various pressure ranges is determined mainly by the vapor pressure of the pump liquid as a function of temperature. Glycerol at normal temperature has a lower vapor pressure than mercury, and at higher temperatures a higher vapor pressure than most commonly used pump oils. The efficient working range of the glycerol vapor pump is therefore wider than that of oil vapor pumps, and they work more efficiently than mercury vapor pumps in the lowest pressure ranges. Glycerol pumps have been tested in continuous use over a long period of time.

Since the first tests, the glycerol vapor pump has passed the experimental stage and has proved its value in practical applications. Two large glycerol vapor pumps with a maximum pumping speed of 1600 l/sec have been in use for more than a year in a continuous process of metal evaporation in which large quantities of paper or Cellophane have been coated with aluminum. The pumps have been in operation continuously day and night while a pressure of a few times 10^{-3} mm of mercury was maintained in a vacuum system of about 10 m^3 capacity. At this pressure, the rate at which gases and vapors were generated during the process and removed by the pumps was of the order of 1500 l/sec. These gases consisted mainly of nitrogen, water vapor and hydrogen. The pumps worked under these conditions satisfactorily, and the original filling of glycerol was not changed for a whole year. Some glycerol escaped, however, through the backing line, and it was therefore necessary to replace the loss after every three months of continuous working.

Glycerol vacuum pumps have also been found suitable for operations in the lowest range of pressures. For this purpose, a very efficient small glass pump has been developed. PAUL ALEXANDER, *J. Sci. Inst.* **25**, 313 (1948).—P. A.

Biological and Psychological Effects of Ultrasonics

Ultrasonic and also sonic waves of high intensity may produce discomfort and even injury, the effects being different at different frequencies. At very low frequencies we find total acceleration, known to aviators as "g." This acceleration is clearly felt as a pressure that forces the aviator down into his seat and causes the familiar sensations at the bottom of a roller coaster dip. Severe prolonging of "g" causes "blackout." At slightly higher frequencies, motion sickness occurs. The organs of equilibrium in the inner ear are involved and visual effects may contribute.

Still higher in the frequency scale, ordinary vibrations, when severe, can be annoying, can impair precision of movement, and can cause serious fatigue. Near the middle of the audible spectrum occurs the resonant frequency of the eyeballs, and in this region, therefore, blurring of vision is likely to be greatest.

In the ultrasonic range, heat effects appear to be the most significant. The very efficient absorption of energy at these frequencies by fur is one reason for their importance. There are also possible dangers to large animals in the ultrasonic region by effects of the same type as cause homogenization of milk, killing of bacteria, and aging of whiskey, but the consequences have not yet been adequately assessed. It may be that man is satisfactorily protected from any such effects of airborne ultrasonics by effective reflection from his skin.

The danger from intense vibration or sound may be injury to a sense organ, particularly to the ear. This may not be as serious as total bodily danger, such as overheating. Small animals may be literally cooked to death by high intensity ultrasonics.

The author suggests and recommends that discomfort and danger levels should be established for the protection of personnel working in high intensity sonic or ultrasonic fields. He suggests that the Ultrasonics Panel of the Aeronautical Board be made a clearing house for the recording of all biological and psychological effects. HALLOWELL DAVIS, *J. Acous. Soc. Am.* **20**, 605 (1948).—T. H. O.

The Place of Mathematics in General Education

It is recognized that there is a great difference between the educational potentials of a subject like mathematics and the realization of these potentials. They cannot be achieved if the subject is taught in such a way that students merely memorize proofs. At some stage in an individual's education, the question must be answered, "What mathematics should be taught, and how?" Properly speaking, there is no one correct answer but many, because human society is not static. What is best suited for one time and place is not for another.

A common difficulty in high schools is to decide how much mathematics should be required of students intending to go to college, and how much of students not intending to go. The answer, in a form flexible enough to meet individual needs, can be given only if there is more cooperation between mathematicians and educators.

There is a considerable amount of evidence to show that it is the student of superior ability rather than the one of limited scholarship who is being seriously neglected in many of our public schools and colleges. To take care of superior students, at least two problems must be solved. One is budgetary. The other involves techniques of selecting superior students from the masses. At the University of North Carolina, and in many other institutions, placement tests are used to sectionize students according to ability in mathematics. They have been found to be very effective. It has also been discovered that students selected solely on the basis of superior grades in a 40-minute mathematics test do distinctly superior work in all other subjects as well.

What kind of mathematics should be taught in high schools cannot be determined apart from the problem of individual differences. Throwing a lot of different topics together in a book does not automatically make that book broadening and a fit instrument for general educational purposes. Work and thought on the part of both author and teacher are required to produce a really satisfactory course. Some students, it is generally recognized, never proceed beyond the particular and the concrete. They cannot learn to generalize and to abstract even in a limited way. It is therefore necessary to make a decision as to when and how the generalizations and abstractions that may be considered characteristic of mathematical thought should be begun. This decision, however, can be made only on the basis of experience, just as it is virtually impossible to determine without actual trial in the classroom whether a particular book or a particular method of presentation will be successful. EDWARD A. CAMERON, *The Mathematics Teacher* **41**, 274 (1948).—T. H. O.

Engineering Education—Plus

The general aim of our engineering schools has been to fit the student to fulfill acceptably the technical requirements of his calling. The formal process for the education of the student engineer has undergone numerous revisions through the more than one hundred years during which engineering education has been under way in the United States. The United States Miliary Academy was, at first, not merely a school to educate men for the career of army officers; it was also a college to supply the young nation's needs for technically trained men, particularly in the field of civil engineering. Later, as a result of expanding industrial plants and the coming of railroads, some of our best-known schools, Rensselaer, Stevens, Cornell and others, came into being.

In reviewing the beginnings of the curricula in these early schools, we find on the one hand an avowed need for those types of instruction which are loosely termed "broadening" and, on the other hand, a singular disregard for the fulfillment of the stated objective. Unreceptive public attitude to the "frills" represented by nonvocational studies and lack of adequate perspective in the faculties themselves limited, for a time, attainment of better balance

between technical subject matter and the arts and humanities. In 1855, B. F. Greene published, in his report as Director of the Rensselaer Polytechnic Institute, a statement: "Of the exceptions, it may be remarked that, in respect to a part—Literature and Philosophy—provisions for regular instruction have been made, but the pressure of the more essential parts of the course has hitherto prevented more than a partial—not proportional—development of these subjects." Two years later Jefferson Davis made the report to President Buchanan that graduates of West Point who were scholars of the first rank in exact sciences were "below mediocrity in polite literature."

In the earliest curricula required for "Mechanical Arts" (Engineering) at Cornell the studies embraced considerable amounts of mathematics, literature, languages, philosophy and history. To these were added lectures by distinguished nonresident professors, among these Louis Agassiz, James Hall, James Russell Lowell and Theodore W. Dwight. Nurtured on such fare, it is small wonder that Cornellians took prominent places in the early "Who's Who" of American engineering. Through the years the national trend toward inclusion of arts and humanities in the engineering curriculum has both risen and fallen. Today, emphasis on the social studies as desirable adjuncts to the education of an engineer is increasing. Obviously, any program which seeks to combine social and technical studies requires complete cooperation between the colleges of engineering and the sister colleges of arts, business and law. Mere fluff dished up as "broadening" cannot long hold the attention of discriminating students. Employers of engineering graduates are displaying increasing interest in the education of nonvocational type, provided it will activate an interest in and understanding of human values.

Quite aside from the attempt to broaden the social outlook of the student is the attempt to inculcate in him an increased realization of the professional status of engineering. Many schools have undertaken program revision to emphasize the meaning and significance of the professional relation. Three requirements upon this program have been set in an address by Dr. Robert E. Doherty: First, a new philosophy and a new outlook which will comprehend the human and social as well as the technical problems; second, a way of thought which faces a problem involving human and social elements as a *whole* problem and not merely a technical problem; third, development of the student's ability to learn from experience so that in the future he can continue to expand his fundamental knowledge, deepen his understanding and improve his power as professional man or women and as citizen.

With the fundamental elements of an engineering education, i.e., science, technology, professionalism and social conscience well established in reasonable proportions, there remains a plus value which the engineering educator may impart by his personal behavior and attitude. This responsibility is to implant in every engineering student a respect for and a desire to possess an "unpledged allegiance to honor." The need of youth is for sane guidance toward the rewards of honest work, charity of viewpoint, tolerance untainted with complacency, a dedication to service—all of those attributes of good citizenship which can be taught by precept and example. The test of the performance of educators of engineers will be the measure of *their* service to mankind.—Presidential address by C. E. MacQuigg at the Austin meeting, American Society for Engineering Education, June 15, 1948. *J. Eng. Educ.* **39**, 11 (1948).—B. H. D.

AMERICAN JOURNAL OF PHYSICS VOLUME 17, NUMBER 1 JANUARY, 1949

ANNOUNCEMENTS AND NEWS

The Upper Atmosphere

During the past decade interest in the upper atmosphere has been rapidly mounting for many reasons. For the first time in history some scientists are able to carry on direct exploration of the higher atmosphere. The V-2 and similar rockets, which reach heights in excess of 100 miles, have carried a wide assortment of instruments for measuring physical conditions in the tenuous air at high levels.

The structure of the upper atmosphere is highly complicated. Scientists have recognized the existence of a number of specific layers, but the composition and nature of these layers change from minute to minute, from season to season, and even from year to year. One of the most significant of these layers, at least from the standpoint of meteorology, is a region that contains a very appreciable amount of ozone. A molecule of ozone, which consists of three atoms of oxygen bound together by chemical forces, strongly absorbs the ultraviolet radiation from the sun. In addition, it cuts out large slices from the solar infra-red. The total energy cut out by the ozone layer tends to heat the surrounding volume of gas. Thus, the layer of maximum ozone content is warmer than any of the regions immediately above or below.

The amount of ozone tends to vary with the season and with solar activity. We are still relatively ignorant concerning the precise character of the equilibria of the ozone layer. Nor do we know what role the ozone plays in the over-all problem of atmospheric circulation. Perhaps studies of the ozone layer will contribute to the solution of the problem of long-range weather forecasting. Certainly the layer acts like a blanket, sending radiation that warms the earth.

Well above the ozone layer lies the ionosphere, which

consists of several layers of electrified gas. Ultraviolet radiation from the sun is again responsible for the electrification. The high-energy radiation pulls electrons away from atoms and molecules in the region. Presumably, each individual ionospheric layer arises from some given chemical constituent of the atmosphere. The lowest layer of the ionosphere, known as the *E* layer, probably comes from the ionization of the oxygen molecule. The two other main layers, the F_1 and F_2, probably arise from ionization of the oxygen atom and the nitrogen molecule, respectively. Scientists have recently suspected the existence of a still higher layer, the *G* layer, the existence of which is tentatively attributed to ionization of nitrogen atoms.

The density of electrification in the various ionospheric layers changes diurnally, seasonally, and annually. Superposed on the more or less regular variations are the disturbances, frequently sudden, caused by solar activity. Knowledge of the nature of the ionospheric layers is extremely important for practical reasons. The ionosphere reflects radio waves around the surface of the earth, making possible long-distance radio communication. Advance knowledge of the occurrence of disturbances is important for the planning of communication schedules, especially since some frequencies are more affected than others by the disturbances.

There are numerous other phenomena of the upper atmosphere which are closely related to solar activity. Among these are the problems of the aurora polaris and of the continuous auroral glow that keeps even the darkest of skies from being perfectly black. These illuminations derive their energy directly from the sun, either from converted ultraviolet radiation or, possibly, from high speed particles ejected from active solar areas.

Thus, the sun becomes an important factor in the understanding of physical conditions in the upper atmosphere. Observations indicate that the 11-year sunspot cycle has associated a variability of many solar features other than spots. These include changes in the activity of prominence explosions, the form and intensity of the solar corona, and —what is perhaps most important from the terrestrial point of view—a marked change in the amount of emitted ultraviolet radiation. New solar programs, to develop improved indices of solar activity, are already under way. These studies should provide information of great benefit to students of atmospheric problems.

Cosmic rays are another interesting and important phenomenon of the upper atmosphere. The origin and precise character of these radiations is not yet completely known. They appear to consist primarily of highly energetic particles, coming in from outer space. Thus, in a sense, they are not phenomena of the upper atmosphere. However, the scientists observe them most effectively in that region, and we have the best chance of observing them in their original state at the highest levels. As the rays descend, the filtering action of air molecules changes the characteristics of the powerful primary rays into secondary particles of lesser energy.

Cosmic rays possess sufficient energy to disrupt atomic nuclei. Thus, scientists consider them one of the primary tools for the study of nuclear forces and reactions. The short-lived mesons, whose masses are intermediate between those of electrons and nuclei, are of special interest.

Because of the fact that primary cosmic rays—some of them at least—possess a positive charge, the magnetic fields of the earth and the sun exert a focusing action upon the radiations. For this reason a redetermination of the magnetic field of the sun is extremely important.

There is even a possibility that changes in the solar magnetic field may, in some way, be responsible for the origin of cosmic rays. However, this recently made suggestion is extremely tentative.

The rapidly accumulating knowledge of conditions in the upper atmosphere will be especially useful at the time—which perhaps is not as far away as the more pessimistic have supposed—when jet or rocket planes may fly their way through the ionosphere. There is a decided acceleration of interest in the problems and information that comes from the indirect studies. Meteors, which are high-speed projectiles from outer space, give valuable data concerning the density, temperature, and pressure in the levels. The echoes of radio signals from the ionospheric layers contribute information of great value. Studies of terrestrial magnetism at high altitudes, measurement of brightness of the sky, and studies of solar radiation in general, all contribute to the knowledge.—DONALD H. MENZEL, AAAS *Symposium on the upper atmosphere*, Washington, D. C., September 15, 1948.

New Members of the Association

The following persons have been made *members* or *junior members* (*J*) of the American Association of Physics Teachers since the publication of the preceding list [Am. J. Physics **16**, 490 (1948)].

Brown, Harry C., Lowell Textile Institute, Lowell, Mass.
Dueker, James Edson, 23 West Madison, Springfield, Ohio.
Earls, David Leigh (*J*), 333 S. Missouri, Liberty, Mo.
Feinberg, Leonard Michael, 562 West 113th St., New York 25, N. Y.
Keane, Robert E. (*J*), 2406 15 Ave. So., Minneapolis 4, Minn.
McIntosh, William Newton, Jr., 929 Westminster St., N.W., Washington, D. C.
McMillan, John G., 1201 So. 52nd St., Omaha, Nebr.
Sullivan, Harris Martin, 1700 Irving Park Rd., Chicago 13, Ill.
Toops, Edward C., Swain Hall, University of Indiana, Bloomington, Ind.